

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 21:57:59 ; Search time 5325 Seconds
(without alignments)
10622.096 Million cell updates/sec

Title: US-10-030-688-1
Perfect score: 1305
Sequence: 1 atggtacctgacagtgatca.....atgtctggaagctgagctg 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:**

- 1: gb_ba:**
- 2: gb_htg:**
- 3: gb_in:**
- 4: gb_om:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vi:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_ov:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vi:**
- 30: em_htg_hum:**
- 31: em_htg_inv:**
- 32: em_htg_other:**
- 33: em_htg_mus:**
- 34: em_htg_pln:**
- 35: em_htg_rod:**
- 36: em_htg_mam:**
- 37: em_htg_vrt:**
- 38: em_sy:**
- 39: em_htgo_hum:**
- 40: em_htgo_mus:**
- 41: em_htgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1305	100.0	1305	6	AX076190	AX076190 Sequence
2	1303.4	99.9	2038	6	AR142620	AR142620 Sequence
3	1303.4	99.9	2038	6	BD137129	BD137129 Human pro
4	1303	99.8	2079	9	BC011703	BC011703 Homo sapi
5	1303	99.8	2081	9	AF179224	AF179224 Homo sapi
6	1298.2	99.5	2137	6	AX207899	AX207899 Sequence
7	1297.4	99.4	1479	6	AX076192	AX076192 Sequence
8	1292	99.0	2070	6	BD235831	BD235831 A novel m
9	1292	99.0	2079	6	AR232520	AR232520 Sequence
10	1292	99.0	2079	6	AX207967	AX207967 Sequence
11	1292	99.0	2079	9	AF216312	AF216312 Homo sapi
12	1263	96.8	2063	6	AX092380	AX092380 Sequence
13	1263	96.8	2063	6	AX376262	AX376262 Sequence
14	1263	96.8	2063	6	AX395214	AX395214 Sequence
15	1263	96.8	2063	6	AX697206	AX697206 Sequence
16	1263	96.8	2063	9	AY358530	AY358530 Homo sapi
17	1207	92.5	1784	9	BC012752	BC012752 Homo sapi
18	922.6	70.7	2268	10	AY043240	AY043240 Mus muscu
19	921	70.6	2060	10	BC021368	BC021368 Mus muscu
20	699	53.6	699	6	E13203	E13203 Human CDNA
21	444.6	34.1	1008	6	BD194909	BD194909 86 human
22	399.2	30.6	609	6	AX076194	AX076194 Sequence
23	312	23.9	395	6	AX070330	AX070330 Sequence
24	194.6	14.9	2992	10	BC042878	BC042878 Mus muscu
25	190.4	14.6	1078	6	AX190545	AX190545 Sequence
26	190.4	14.6	1314	6	AX190543	AX190543 Sequence
27	190.4	14.6	1314	6	AX190546	AX190546 Sequence
28	190.4	14.6	1614	6	AX360089	AX360089 Sequence
29	190.4	14.6	1689	6	AX149579	AX149579 Sequence
30	190.4	14.6	1748	6	AX369323	AX369323 Sequence
31	190.4	14.6	3222	9	BC018715	BC018715 Homo sapi
32	190.4	14.6	3324	9	AB048797	AB048797 Homo sapi
33	189.6	14.5	1222	6	BD276278	BD276278 Serine pr
34	189.6	14.5	2067	6	AX149581	AX149581 Sequence
35	189.6	14.5	2237	9	AY190317	AY190317 Homo sapi
36	189.6	14.5	2393	9	AB048796	AB048796 Homo sapi
37	175.8	13.5	2752	10	MMU300738	AJ300738 Mus muscu
38	171.4	13.1	176526	2	AP001192	AP001192 Homo sapi
39	171.4	13.1	215441	9	AP002800	AP002800 Homo sapi
40	165.8	12.7	176526	2	AP001192	AP001192 Homo sapi
41	164.6	12.6	1365	6	AX360088	AX360088 Sequence
42	164.6	12.6	1394	6	AX392866	AX392866 Sequence
43	164.6	12.6	1394	6	BD276279	BD276279 Serine pr
44	164.6	12.6	2135	6	AX277411	AX277411 Sequence
45	164.6	12.6	2135	9	AB038159	AB038159 Homo sapi

ALIGNMENTS

RESULT 1
AX076190
LOCUS AX076190 1305 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 1 from Patent WO0104141.
ACCESSION AX076190
VERSION AX076190.1 GI:12710815
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Suendermann,B., Hofmann,U., Matzku,S. and Wilbert,O.
TITLE Seripancrin
JOURNAL Patent: WO 0104141-A 1 18-JAN-2001;

MERCK PATENT GmbH (DE)
FEATURES
Location/Qualifiers
1. 1305
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
1. >1305
/note="unnamed protein product"
CDS
/codon_start=1
/protein_id="CAC28479.1"
/db_xref="GI:12710816"
/db_xref="REMBL:CAC28479"
/translation="MDPSDQPLNSLDVKLRPRIPMETFRKVGPIIILSLASI
IIIVVLKILDKYFLCGPLHIFPRKQLCDGELCPGLEDHEHCVKSPPEGPFAV
RLSDRLTQVLDSTGNWFSACFDNFTEALAEATACRQMGYSKPTFRVAFVIGPDOL
DVVEITNSBELRMNSGSLGSLVSLHCLACKSLKTPRVVGGEEASVDSWPMQV
SIQYDKHVGSGSLDHPHWLTAHCFKHTDFVNMKVRAGSKLGSFPLAVAKII
IEFNPMYPKNDIALMKLQPLPFTSFTVPICLPFEDELTPATPLKILGWGFTKNG
KMSDILQASVQVDSIRCNADDAIOGEVTEKMKCAGIEGGVDTCQDSSGSLMIVQ
SDQHWVGVIVSWGYGCGPSTPGVYTKVSAYLNIWNIYNWKAEL"
ORIGIN
Query Match 100.0%; Score 1305; DB 6; Length 1305;
Best Local Similarity 100.0%; Pred. No. 7.5e-302;
Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGATCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGGCGAAACCC 60
DB 1 ATGGATCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGGCGAAACCC 60
QY 61 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 61 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120
QY 121 CTGGCGAGTATCATATGTGTTGTCTCTATCAAGGTGATTTGGATAAATACTACTTC 180
DB 121 CTGGCGAGTATCATATGTGTTGTCTCTATCAAGGTGATTTGGATAAATACTACTTC 180
QY 181 CTCTGGCGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGACGGAGAGCTGGAC 240
DB 181 CTCTGGCGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGACGGAGAGCTGGAC 240
QY 241 TGTCTCTGGCGAGGAGCAGGAGCAGCTGTGACAGAGCTTCCCGAAGGCGCTGCAAGTG 300
DB 241 TGTCTCTGGCGAGGAGCAGGAGCAGCTGTGACAGAGCTTCCCGAAGGCGCTGCAAGTG 300
QY 301 GCAGTCCGCTCTCAAGGACCGATCAACACTGAGGTCTGGAATCGGACACAGGGAAC 360
DB 301 GCAGTCCGCTCTCAAGGACCGATCAACACTGAGGTCTGGAATCGGACACAGGGAAC 360
QY 361 TGGTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAG 420
DB 361 TGGTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTAGGCAG 420
QY 421 ATGGCTACAGACAGAAACCCACTTTGAGACTGTGAGATTTGCCAGACCGAGATCTG 480
DB 421 ATGGCTACAGACAGAAACCCACTTTGAGACTGTGAGATTTGCCAGACCGAGATCTG 480
QY 481 GATGTGTTGAATACAGAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCC 540
DB 481 GATGTGTTGAATACAGAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCC 540
QY 541 TGTCTCTCAGGCTCCCTGGTCTCCCTGACATGTTCTGCTGTGGGAAGAGCCTGAAGAC 600
DB 541 TGTCTCTCAGGCTCCCTGGTCTCCCTGACATGTTCTGCTGTGGGAAGAGCCTGAAGAC 600
QY 601 CCCGCTGTGTGGTGGGAGAGGCTCTGTGATTTGGCTTGGCTGGCAGGTCAGCATC 660
DB 601 CCCGCTGTGTGGTGGGAGAGGCTCTGTGATTTGGCTTGGCTGGCAGGTCAGCATC 660
QY 661 CAGTACGACAAACAGCAGCTGTGGAGGGAGCATCTCTGGACCCCACTGGGCTCTCAG 720
DB 661 CAGTACGACAAACAGCAGCTGTGGAGGGAGCATCTCTGGACCCCACTGGGCTCTCAG 720

QY 721 GCAGCCCACTGCTTCAGGAAACATACCGATGTTTCAACTGGAAGGTGCGGCGAGGTCA 780
DB 721 GCAGCCCACTGCTTCAGGAAACATACCGATGTTTCAACTGGAAGGTGCGGCGAGGTCA 780
QY 781 GACAAACTGGGAGCTTCCCATCCCTGGCTGGGCGAGATCATCATATTGAATTCAC 840
DB 781 GACAAACTGGGAGCTTCCCATCCCTGGCTGGGCGAGATCATCATATTGAATTCAC 840
QY 841 CCATGTATACCCCAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTACTTC 900
DB 841 CCATGTATACCCCAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTACTTC 900
QY 901 TCAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCAC 960
DB 901 TCAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCAC 960
QY 961 CCATCTGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTTGACATA 1020
DB 961 CCATCTGATCATTTGGATGGGCTTTACGAAGCAGATGGAGGGAAGATGTTGACATA 1020
QY 1021 CTGCTGCAGGCTCAGTCCAGTCAATGACACACACCGTGCATGACAGATCGGTAC 1080
DB 1021 CTGCTGCAGGCTCAGTCCAGTCAATGACACACACCGTGCATGACAGATCGGTAC 1080
QY 1081 CAGGGGAGTCAACGAGAGATGATGTGTCAGGATCCCGGAGGGGTGTGACACC 1140
DB 1081 CAGGGGAGTCAACGAGAGATGATGTGTCAGGATCCCGGAGGGGTGTGACACC 1140
QY 1141 TGCCAGGCTGACAGTGGTGGGCTTGTATCAATCTGACCACTGTCATGTGGTGGC 1200
DB 1141 TGCCAGGCTGACAGTGGTGGGCTTGTATCAATCTGACCACTGTCATGTGGTGGC 1200
QY 1201 ATCGTGTAGTGGGCTATGCTGCGGGGCGCCGAGCACCAGGAGTATACCAAGGTC 1260
DB 1201 ATCGTGTAGTGGGCTATGCTGCGGGGCGCCGAGCACCAGGAGTATACCAAGGTC 1260
QY 1261 TCAGGCTATCTCAACTGGATCTTACAATGTCTGGAAGGCTGAGCTG 1305
DB 1261 TCAGGCTATCTCAACTGGATCTTACAATGTCTGGAAGGCTGAGCTG 1305
RESULT 2
AR142620
LOCUS AR142620 2038 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 18 from patent US 6203979.
ACCESSION AR142620
VERSION AR142620.1 GI:15103906
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2038)
AUTHORS Bandman, O., Hillman, J. L., Yue, H., Guegler, K. J., Corley, N. C.,
Tang, Y. I. M., and Shan, P.
TITLE Human protease molecules
JOURNAL Patent: US 6203979-A 18 20-MAR-2001;
FEATURES
source
1. 2038
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 99.9%; Score 1303.4; DB 6; Length 2038;
Best Local Similarity 99.9%; Pred. No. 1.8e-301;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGATCTGACAGTATCAACCTCTGAAACAGCTCGATGTCAAAACCCCTGGCGAAACCC 60
DB 200 ATGGATCTGACAGTATCAACCTCTGAAACAGCTCGATGTCAAAACCCCTGGCGAAACCC 259
QY 61 CGTATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120

```
Db 260 CGTATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCCATCATCATAGCACTACTGAGC 319
Qy 121 CTGGCAGATCATCATTTGTGTTGTTCTCTCATCAAGTGATTTGGGATAAATACTACTTC 180
Db 320 CTGGCAGATCATCATTTGTGTTGTTCTCTCATCAAGTGATTTGGGATAAATACTACTTC 379
Qy 181 CTCTGGGGCAGCTCTCCACTTCATCCCGAGGAGAGAGCTGTGTGACGGAGAGCTGGAC 240
Db 380 CTCTGGGGCAGCTCTCCACTTCATCCCGAGGAGAGAGCTGTGTGACGGAGAGCTGGAC 439
Qy 241 TGTCCCTTTGGGGAGGAGGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 300
Db 440 TGTCCCTTTGGGGAGGAGGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 499
Qy 301 GCAGTCCGCTCTCCAGGACCGATCCACATCGCAGGTGTGTGAGCTCGGGCCACAGGGAAC 360
Db 500 GCAGTCCGCTCTCCAGGACCGATCCACATCGCAGGTGTGTGAGCTCGGGCCACAGGGAAC 559
Qy 361 TGGTCTCTGCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTGAGGAG 420
Db 560 TGGTCTCTGCTGCTGTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTGAGGAG 619
Qy 421 ATGGGCTACAGAGCAAAACCCACTTTCAGAGCTGTGAGAGATTGCCCCAGACAGGATCTG 480
Db 620 ATGGGCTACAGAGCAAAACCCACTTTCAGAGCTGTGAGAGATTGCCCCAGACAGGATCTG 679
Qy 481 GATGTTGTTGAATCACAGAAACAGCAGAGCTTCGCATGCGGAACTCAAGTGGGCC 540
Db 680 GATGTTGTTGAATCACAGAAACAGCAGAGCTTCGCATGCGGAACTCAAGTGGGCC 739
Qy 541 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAGAGGCTGAAGACC 600
Db 740 TGTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAGAGGCTGAAGACC 799
Qy 601 CCCGCTGTGGTGGGAGAGGAGGCTCTGTGATTTCTGGCTTGGGCTTGGCAGGTGAGATC 660
Db 800 CCCGCTGTGGTGGGAGAGGAGGCTCTGTGATTTCTGGCTTGGGCTTGGCAGGTGAGATC 859
Qy 661 CAGTACGACAAACAGCAGCTCTGTGGAGGAGAGCTCTGACCCCTCACTGGGTCTCAAG 720
Db 860 CAGTACGACAAACAGCAGCTCTGTGGAGGAGAGCTCTGACCCCTCACTGGGTCTCAAG 919
Qy 721 GCAGCCCACTGTTCCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCA 780
Db 920 GCAGCCCACTGTTCCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGGCAGGCTCA 979
Qy 781 GACAACTGGCAGCTTCCCATCCCTGCTGTGCGCAGATCATCATGATTAATTAAC 840
Db 980 GACAACTGGCAGCTTCCCATCCCTGCTGTGCGCAGATCATCATGATTAATTAAC 1039
Qy 841 CCCATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAAGTTCCTCCACTCACTTTC 900
Db 1040 CCCATGTACCCCAAGACATGACATCGCCCTCATGAAGCTGCAAGTTCCTCCACTCACTTTC 1099
Qy 901 TCAGGCAAGTTCAGGCCCATCTGTGTGCTGCTTCTTGTGATGAGAGCTCATCTCCAGCACC 960
Db 1100 TCAGGCAAGTTCAGGCCCATCTGTGTGCTGCTTCTTGTGATGAGAGCTCATCTCCAGCACC 1159
Qy 961 CCACCTCTGGATCAATGGATGGGGCTTTACGAAGCAGAAATGAGGGAAGATGTTGACATA 1020
Db 1160 CCACCTCTGGATCAATGGATGGGGCTTTACGAAGCAGAAATGAGGGAAGATGTTGACATA 1219
Qy 1021 CTGTGACAGGCTGAGTCCAGGTCAATTGACAGCACACGGTGCATGACAGAGATGCGTAC 1080
Db 1220 CTGTGACAGGCTGAGTCCAGGTCAATTGACAGCACACGGTGCATGACAGAGATGCGTAC 1279
Qy 1081 CAGGGGGAAGTCAACGAGAAGATGATGTGTCAGGCATCCCGAAGGGGTGTGACACC 1140
Db 1280 CAGGGGGAAGTCAACGAGAAGATGATGTGTCAGGCATCCCGAAGGGGTGTGACACC 1339
Qy 1141 TGCCAGGCTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGC 1200
Db 1340 TGCCAGGCTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGGC 1399
```

```
Qy 1201 ATCGTTAGCTGGGGCTATGGCTCGGGGGCCCGAGACCCCGAGAGTATACACCAAGGTC 1260
Db 1400 ATCGTTAGCTGGGGCTATGGCTCGGGGGCCCGAGACCCCGAGAGTATACACCAAGGTC 1459
Qy 1261 TCAGCCCTATCTCAACCTGATCTACATGCTGAAAGGCTGAGCTG 1305
Db 1460 TCAGCCCTATCTCAACCTGATCTACATGCTGAAAGGCTGAGCTG 1504

RESULT 3
BD137129
LOCUS BD137129 2038 bp DNA linear PAT 18-SEP-2002
DEFINITION Human protease molecule.
ACCESSION BD137129
VERSION BD137129.1 GI:23232074
KEYWORDS JP 2002508970-A/6.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2038)
AUTHORS Bandman,O., Hillman,J.L., Yue,H., Guegler,K.J., Corley,N.C.,
Tang,T.Y. and Shah,P.
TITLE Human protease molecule
JOURNAL Patent: JP 2002508970-A 6 26-MAR-2002;
INCYTE PHARMACEUTICALS INC
OS Homo sapiens (human)
PN JP 2002508970-A/6
PD 26-MAR-2002
PF 12-JAN-1999 JP 2000540252
PR 16-JAN-1998 US 09/008271
PI OLGA BANDMAN,JENNIFER L HILLMAN,HENRY YUE,KARL J GUEGLER,NEIL
PI C CORLEY,
PI TOM Y TANG,PURVI SHAH
PC C12N15/09,A61K38/46,C07K16/40,C12N1/19,C12N1/21,C12N5/10 PC
PC C12N9/48,C12N5/64,
PC C12Q1/68,C12N15/00,A61K37/54,C12N5/00
CC 1337018
FH Key Location/Qualifiers
FT source 1..2038
FT Location/Qualifiers
1..2038
/organism='Homo sapiens (human)'
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

FEATURES
source
1..2038
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'

ORIGIN
Query Match 99.9%; Score 1303.4; DB 6; Length 2038;
Best Local Similarity 99.9%; Pred. No. 1.8e-301;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGATCCTGACAGTATCAACCTCTGAAACAGGCTCGATGTCAAAACCCCTGGCGAAACCC 60
Db 200 ATGGATCCTGACAGTATCAACCTCTGAAACAGGCTCGATGTCAAAACCCCTGGCGAAACCC 259
Qy 61 CGTATCCCATGAGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
Db 260 CGTATCCCATGAGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 319
Qy 121 CTGGCAGATCATCATATTGTGTTGTTCTCTCATCAAGTGATTTCTGGATAAATACTACTTC 180
Db 320 CTGGCAGATCATCATATTGTGTTGTTCTCTCATCAAGTGATTTCTGGATAAATACTACTTC 379
Qy 181 CTCTGGGGCAGCTCTCCACTTCATCCCGAGGAGAGAGCTGTGTGACGGAGAGCTGGAC 240
Db 380 CTCTGGGGCAGCTCTCCACTTCATCCCGAGGAGAGAGCTGTGTGACGGAGAGCTGGAC 439
Qy 241 TGTCCCTTTGGGGAGGAGGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 300
Db 440 TGTCCCTTTGGGGAGGAGGAGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTGCACTG 499
```


	/lab_host="DH10B-R"
	/note="vector: pOTB7"
gene	1. .2079
	/genes="TMPRSS4"
	/notes="synonyms: TMPRSS3, MT-SP2"
	/db_xref="LocusID:56643"
	/db_xref="MIM:606585"
CDS	208..1521
	/codon_start=1
	/product="transmembrane protease, serine 4, isoform 1"
	/protein_id="AAH11703.1"
	/db_xref="GI:15079795"
	/db_xref="LocusID:56643"
	/translations="MIQDPDSDPLNSLDVKPKRIPDMETFRKVGIFIIALLSLA SIIIVVLIRIYDIKYLFCGPHFIPRKQLCDGLDCPLGEDHCWKVSPPEPAV AVLSKDRSTQLVDATGNWGFACDFNFTALAEATACRMGYSSKPTFAVEIGPDQ DLDVBITENSOLBRWNSSGPLSGSLVSLHLCLAGCKLIKTPRVVGGEEASVDSWP ITIEFNPMYPKDNDILMDPHWLTAACHCFKRKTDFVNFKVRAGSKLGSPPSLAVALKI IITEFNPMYPKDNDILMDPHWLTFSTGVTRPILFFDEBELTPATPLMTIGWGFTKO NGKSGDILLQASQVVIDSTRCNADDAYQGEVTEKMCAGIPESGVDTCCGDSGGFLM YQSQRHVIVGSVSWGCGGPSFGVTVKVSAYLINMYNWRAEL"
misc_feature	379..483
	/note="LDua; Region: Low Density Lipoprotein Receptor Class A domain, a cysteine-rich repeat that plays a central role in mammalian cholesterol metabolism"
	/db_xref="CDD:cd00112"
misc_feature	517..789
	/note="SR; Region: Scavenger receptor Cys-rich"
	/db_xref="CDD:smart00202"
misc_feature	817..1494
	/note="Tryp SPc; Region: Trypsin-like serine protease"
	/db_xref="CDD:smart0020"
ORIGIN	
Query Match	99.8%; Score 1303; DB 9; Length 2079;
Best Local Similarity	100.0%; Pred. NO. 2.2e-301;
Matches 1303; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	3 GGATCTGTGACAGTGATCAAACCTCTGAACAGCTTCGATGTCAAACCCCCTGGCAAAACCCCCG 62
Dd	216 GGATCTGTGACAGTGATCAAACCTCTGAACAGCTTCGATGTCAAACCCCCTGGCAAAACCCCCG 275
Qy	63 TATCCCCCATGGAGACTTTCAGAAAGTTGGGATGCCCATCATCATAGCACTACTAGACCT 122
Dd	276 TATCCCCCATGGAGACTTTCAGAAAGTTGGGATGCCCATCATCATAGCACTACTAGACCT 335
Qy	123 GGGAGGTATCATATTGTGTGTTCCTCATCAGGTGATTCGATAAATACTACTTCTCT 182
Dd	336 GGGAGGTATCATATTGTGTGTGTTCCTCATCAGGTGATTCGATAAATACTACTTCTCT 395
Qy	183 CTGGGGCACGCCTCTCCAATTTCATCCGAGGAAGACAGCTGTGTGACGGAGAGCTGGACTG 242
Dd	396 CTGGGGCACGCCTCTCCAATTTCATCCGAGGAAGACAGCTGTGTGACGGAGAGCTGGACTG 455
Qy	243 TCCTTTGGGGAGACAGAGAGCACTGTGTGACAGAGCTTCCCGAAGGCGCTCGAGTGGC 302
Dd	456 TCCTTTGGGGAGACAGAGAGCACTGTGTGACAGAGCTTCCCGAAGGCGCTCGAGTGGC 515
Qy	303 AGTCGGGCTCTTCAAAGAACCGATCCACACTGACAGTGTGATTCGAGTTCGAGGAGTGGGAACTG 362
Dd	516 AGTCGGGCTCTTCAAAGAACCGATCCACACTGACAGTGTGATTCGAGTTCGAGGAGTGGGAACTG 575
Qy	363 GTTCTGTGCTGTGTTTGCAGAACTTTCAGAAAGCTTCTCGCTGACAGACCTGTGAGGCAGAT 422
Dd	576 GTTCTGTGCTGTGTTTGCAGAACTTTCAGAAAGCTTCTCGCTGACAGACCTGTGAGGCAGAT 635
Qy	423 GGGCTACAGACAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACACAGGATCTGGA 482
Dd	636 GGGCTACAGACAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACACAGGATCTGGA 695
Qy	483 TGTTGTTGAAATCACAGAAAACACCCAGGAGCTTCGCATCGGAACTCAAGTGGGCGCTG 542
Dd	696 TGTTGTTGAAATCACAGAAAACACCCAGGAGCTTCGCATCGGAACTCAAGTGGGCGCTG 755

PUBMED 10825129
 REFERENCE 2 (bases 1 to 2081)
 AUTHORS Wallrapp, C. and Gress, T.M.
 TITLE Direct Submission
 JOURNAL Submitted (20-AUG-1999) Internal Medicine I, University of Ulm,
 Robert-Koch-Street 8, Ulm, Baden-Wuerttemberg 89081, Germany
 FEATURES
 Location/Qualifiers
 1..2081
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q23.3"
 /tissue_type="pancreatic carcinoma"
 1..2081
 /gene="TMPRSS3"
 215..1528
 /gene="TMPRSS3"
 /codon_start=1
 /product="transmembrane serine protease 3"
 /protein_id="AAF74526.1"
 /db_xref="GI:8347149"
 /translation="MLQDPDSQPLNSLDVPLKPRIPMETFRKVGIPILIIALLSLA
 SIIIVVLIKVIDLYFLOQPLHFIPIKQLCDGELDCPLGDEEHCVKSFPEGPAV
 AVRLSKRSTIQEVLDSATGNWFSACFDNFTEALAEATACRQVGYSKPTFRAVIGPDQ
 DLIVVEITENSQELMRNSGPKLSGLSLVSLHCLACGKSLKTRPVVGGEASVDSWFW
 QVSTQYDKOHVCGGSLIDPHWLTAACHFRKHTDFVNMKVRAGSDKLGSFPLAVAKI
 IIEFNPMYKNDIALMKLOPPLTFSTGVRPCLPFPDEBELPATPLWIIWGFTKQ
 NGKXSDILLOASVQVDSIRCNADDAVQSEVTEKMCAGIPSGGVDTQGDSDSGPLM
 YQSDQHVGVGWSVGGCGSPTEGVITKVSAYLNWLNWKAEL"
 215..307
 /gene="TMPRSS3"
 /note="Region: cytoplasmic domain"
 308..373
 /gene="TMPRSS3"
 /note="Region: transmembrane domain"
 374..1525
 /gene="TMPRSS3"
 /note="Region: extracellular domain"
 602..604
 /gene="TMPRSS3"
 /note="glycosylation site"
 746..748
 /gene="TMPRSS3"
 /note="glycosylation site"
 /note="glycosylation site"
 order(800..802,1142..1144)
 /gene="TMPRSS3"
 /note="disulfide bond"
 order(902..904,950..952)
 /gene="TMPRSS3"
 /note="disulfide bond"
 947..949
 /gene="TMPRSS3"
 /note="active site"
 1082..1084
 /gene="TMPRSS3"
 /note="active site"
 /note="active site"
 order(1280..1282,1328..1330)
 /gene="TMPRSS3"
 /note="disulfide bond"
 1373..1375
 /gene="TMPRSS3"
 /note="active site"
 ORIGIN
 Query Match 99.8%; Score 1303; DB 9; Length 2081;
 Best Local Similarity 100.0%; Pred. No. 2.2e-301;
 Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GGATCCCTGACAGTGCATCAACCTCTGACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 62
 DB 223 GGATCCCTGACAGTGCATCAACCTCTGACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 282

QY 63 TATCCCATGGAGACCTTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
 DB 283 TATCCCATGGAGACCTTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 342
 QY 123 GCGAGTATCATCAITGTGTTGTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 182
 DB 343 GCGAGTATCATCAITGTGTTGTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 402
 QY 183 CTGGGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 242
 DB 403 CTGGGGGAGCCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACTG 462
 QY 243 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTTGCAGTGGC 302
 DB 463 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTTGCAGTGGC 522
 QY 303 AGTCCGCTCTCCAGGACCGATCCACACTGAGAGTGTGTGAGCTGGGCGACAGGGAATG 362
 DB 523 AGTCCGCTCTCCAGGACCGATCCACACTGAGAGTGTGTGAGCTGGGCGACAGGGAATG 582
 QY 363 GTTCTCTGCTCTTTCGACAACTTCAAGAGCTCTCGCTGAGAGAGCTGTAGGACAGT 422
 DB 583 GTTCTCTGCTCTTTCGACAACTTCAAGAGCTCTCGCTGAGAGAGCTGTAGGACAGT 642
 QY 423 GGGCTACAGCAACAAACCACTTTTCAAGAGTGTGTGAGATTTGGCCAGACAGGATCTTGA 482
 DB 643 GGGCTACAGCAACAAACCACTTTTCAAGAGTGTGTGAGATTTGGCCAGACAGGATCTTGA 702
 QY 483 TGTGTGTAATTCAGAAACAGCAGGAGCTTCCGATCGGAGTCAAGTGGGCCCTG 542
 DB 703 TGTGTGTAATTCAGAAACAGCAGGAGCTTCCGATCGGAGTCAAGTGGGCCCTG 762
 QY 543 TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTCCCTGTGGGAGAGCCTGAAGACCCC 602
 DB 763 TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTCCCTGTGGGAGAGCCTGAAGACCCC 822
 QY 603 CCGTGTGGTGGTGGGAGGAGGCTCTGTGAGTCTTTGGGCTTTGGCAGGTGAGGATCCA 662
 DB 823 CCGTGTGGTGGTGGGAGGAGGCTCTGTGAGTCTTTGGGCTTTGGCAGGTGAGGATCCA 882
 QY 663 GTACGACAAACAGCAGCTGTGTGAGGAGGAGTCTTGGACCCCACTGGTCTTCCAGGCG 722
 DB 883 GTACGACAAACAGCAGCTGTGTGAGGAGGAGTCTTGGACCCCACTGGTCTTCCAGGCG 942
 QY 723 AGCCCACTGCTTTCAGAAACATACCGATGTGTAACTGGAAGGTGCGGGGAGGCTCAGA 782
 DB 943 AGCCCACTGCTTTCAGAAACATACCGATGTGTAACTGGAAGGTGCGGGGAGGCTCAGA 1002
 QY 783 CAAACTGGGACGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACACCC 842
 DB 1003 CAAACTGGGACGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCACACCC 1062
 QY 843 CATGTACCCCAAGAACATATGACATCGCCCTCATGAAGCTGAGTTCCCACTCACTTTCTC 902
 DB 1063 CATGTACCCCAAGAACATATGACATCGCCCTCATGAAGCTGAGTTCCCACTCACTTTCTC 1122
 QY 903 AGGCACAGTCAAGGCCCTCTGTCTCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 962
 DB 1123 AGGCACAGTCAAGGCCCTCTGTCTCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC 1182
 QY 963 ACTCTGGATCATTTGAGTGGGCTTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1022
 DB 1183 ACTCTGGATCATTTGAGTGGGCTTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1242
 QY 1023 GGTGAGGCTCACTCCAGGTCAATTCAGACACAGGTGCAATGCAGACGATGGGTACCA 1082
 DB 1243 GGTGAGGCTCACTCCAGGTCAATTCAGACACAGGTGCAATGCAGACGATGGGTACCA 1302
 QY 1083 GGGGGAAGTCAACCGAAGATGATGTGTGAGGCAATCCCGAAGGGGGTGTGGACACCTG 1142
 DB 1303 GGGGGAAGTCAACCGAAGATGATGTGTGAGGCAATCCCGAAGGGGGTGTGGACACCTG 1362
 QY 1143 CCAGGCTGACAGTGTGGGCCCCCTGATGTACCACTCTGACACGATGGTGGGAT 1202

Db 1363 CCAGGGTGACAGTGGTGGGCCCTTGATGATACCAATCTGACAGTGGCATGTGGGTCAT 1422
Qy 1203 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGAGTATACACCAAGGTCTC 1262
Db 1423 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGAGTATACACCAAGGTCTC 1482
Qy 1263 AGCCTATCTCAACTGGATCTACAACTGCTGGAAGGCTGAGCTG 1305
Db 1483 AGCCTATCTCAACTGGATCTACAACTGCTGGAAGGCTGAGCTG 1525

RESULT 6
AX207899 2137 bp DNA linear PAT 31-AUG-2001
LOCUS
DEFINITION Sequence 3 from Patent WO0157194.
ACCESSION AX207899
VERSION AX207899.1 GI:15422497
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE Nucleic acid molecules encoding transmembrane serine proteases, the
JOURNAL encoded proteins and methods based thereon
PATENT: WO 0157194-A 3 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
source
1. 2137
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
251..1574
/note="unnamed protein product; DNA sequence encoding a
transmembrane serine protease (MtsP3) protein"
/codon_start=1
/protein_id="CAC60378.1"
/db_xref="GI:15422498"
/translation="MLQPPSDQPLNSLDVPLKPKRIPMETRKRKVPILIIALLSLA
SIITVVLRIKILVDYFLGQPLHIFPRKOLCDGELDCPLGEDEEHCVCPEFGFAV
AVRLSKDSTLQVLDSDATNMFSAFDNFTAEALACRQMGYSKFTFRAVIGPDQ
DLSVEITENSOBLMENSQPLSLGLSLHCLACGSLKTRPVVKGEEASVSWPW
QVSTQYDIHQVCGSILDPHVLTAACHFRKHTDVNNKVRAGSKLGSFPLAVAKI
IIIFENPMYKDNIDIALMKLOPLTFSTVLELCLPFDEBELPATPLWIIIGGFTQ
NGKMSDILLOASQVVIDSRNADDAVQGEVTEKMKCAGIPGGVDTCCGDSGGPLM
YQSDQHVHVGIVSGYCGGPESTPEVTIKVSAYLNWYNVRAEL"

ORIGIN
Query Match 99.5%; Score 1298.2; DB 6; Length 2137;
Best Local Similarity 99.8%; Pred. No. 3.1e-300;
Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 GGATCCTGACAGTATCAACTCTGAAACAGCCTGATGTCATAACCCCTGCGCAACCCCG 62
Db 269 GGATCCTGACAGTATCAACTCTGAAACAGCCTGATGTCATAACCCCTGCGCAACCCCG 328
Qy 63 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 329 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCCT 388
Qy 123 GCGAGTATCATATGTTGGTGTCTCATCAAGGTGATCTGGAATAATACTACTTCTCT 182
Db 389 GCGAGTATCATATGTTGGTGTCTCATCAAGGTGATCTGGAATAATACTACTTCTCT 448
Qy 183 CTGCGGGAGCCTCTCCACTTCATCCGAGAGCAGCTGTGACCGAGAGCTGAGCTG 242
Db 449 CTGCGGGAGCCTCTCCACTTCATCCGAGAGCAGCTGTGACCGAGAGCTGAGCTG 508
Qy 243 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCTGCAAGTGGC 302

RESULT 7
AX076192

DIADEXUS INC					
OS	Homo sapiens (human)				
FN	JP 2002523760-A/2				
PD	30-JUL-2002				
PF	01-SEP-1999 JP 2000567741				
PR	02-SEP-1998 US 60/098880				
PI	SUSANA SALCEDA, XONGMING SUN, HERVE RECIPON, ROBERT CAFFERKEY PC				
Goin33/574,	A61K39/395, A61K39/395, A61K39/395, A61K39/395 PC				
, A61K49/00,	A61K49/00,				
PC	A61K51/00, C07K16/32, C12N15/09, C12Q1/68, C12N15/00, A61K49/02 CC				
A novel method of diagnosing, monitoring, staging, imaging and treating					
CC various cancers					
FH key	Location/Qualifiers				
FT source	1..2070				
FEATURES	Location/Qualifiers				
source	1..2070				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
ORIGIN					
Query Match	99.0%; Score 1292; DB 6; Length 2070;				
Best Local Similarity	99.9%; Pred. No. 9,5e-299;				
Matches 1303; Conservative	0; Mismatches 0; Indels 1; Gaps 1;				
Qy	3 GGATCCTTGACAGTGCATCAACTCTGAAACAGCGCTGGATGTCAAACCCCTTCGCGAAACC	62			
Db	222 GGATCCTTGACAGTGCATCAACCTCTGAACAGCGCTGGATGTCAAACCCCTTCGCGAAACC	281			
Qy	63 TATCCCATTGAGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACCTACTGAGCCT	122			
Db	282 TATCCCATTGAGAGACCTTCAGAAAGGTGGGATCCCCATCATCATAGCACCTACTGAGCCT	341			
Qy	123 GCGAGGTATCATCATTTGTGGTTGTCCTCATCAAGGTGATCTTGGATAAATACTACTTCT	182			
Db	342 GCGAGGTATCATCATTTGTGGTTGTCCTCATCAAGGTGATCTTGGATAAATACTACTTCT	401			
Qy	183 CTGGGGCGAGCCTCTCCACTTTCATCCCGAGAACGCGTGTGTGACGGAGAGCTGGACTG	242			
Db	402 CTGGGGCGAGCCTCTCCACTTTCATCCCGAGAGCAGCTGTGTGACGGAGAGCTGGACTG	461			
Qy	243 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCAGAGGCCCTGCAGTGGC	302			
Db	462 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCAGAGGCCCTGCAGTGGC	521			
Qy	303 AGTCGGCTCTCCAAGGACCGATCCACACTGCGAGTGTGAGCTCGGCCACAGGGAAC	362			
Db	522 AGTCGGCTCTCCAAGGACCGATCCACACTGCGAGTGTGAGCTCGGCCACAGGGAAC	581			
Qy	363 GTTCTCTGCGCTGTTTCGACAACCTTCAGAAAGCTCTCGCTCAGACAGCGCTCTAGGCGAG	422			
Db	582 GTTCTCTGCGCTGTTTCGACAACCTTCAGAAAGCTCTCGCTCAGACAGCGCTCTAGGCGAG	641			
Qy	423 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACACAGGATCTGGA	482			
Db	642 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACACAGGATCTGGA	701			
Qy	483 TTGTTTGAATACAGAAAAACGACGAGGCTTTCGATCGGAACTCAAGTGGGCCCTG	542			
Db	702 TTGTTTGAATACAGAAAAACGACGAGGCTTTCGATCGGAACTCAAGTGGGCCCTG	761			
Qy	543 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGCTTTCCTGTGGGAAGAGCCTGAAGACCC	602			
Db	762 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGCTTTCCTGTGGGAAGAGCCTGAAGACCC	821			
Qy	603 CCGTGTGGTGGGGAGGAGGCGCTCTGTGGAATCTTGGCCCTTGGAGCTCAGCATCCA	662			
Db	822 CCGTGTGGTGGGGAGGAGGCGCTCTGTGGAATCTTGGCCCTTGGAGCTCAGCATCCA	881			
Qy	663 GTACGAAAACAGCAGCGCTCTGTGAGGAGGAGCATCCTCGAACCCCACTGGGTCCTCAC-GG	721			

D	b	882	GTACGACAAACAGACACCTCTGTGGAGGAGCATCCTGGACCCCACTGGGTCTTCACGGG	941
Q	y	722	CAGCCCCTGCTTCCAGGAACATACCAGATGTGTTCAACTGGAAGGTGCGGCAGGCTCAG	781
D	b	942	CAGCCCCTGCTTCCAGGAACATACCAGATGTGTTCAACTGGAAGGTGCGGCAGGCTCAG	1001
Q	y	782	ACAACTGGGAGCTTCCCATCCCTGGCTGTGGCAAAGATCATCATATTGAATCAACC	841
D	b	1002	ACAACTGGGAGCTTCCCATCCCTGGCTGTGGCAAAGATCATCATATTGAATCAACC	1061
Q	y	842	CCATGTACCCCAAAGAACAATGACATGCCCTCATGAAGCTGCAGTTCCTCACTCACTTTCT	901
D	b	1062	CCATGTACCCCAAAGAACAATGACATGCCCTCATGAAGCTGCAGTTCCTCACTCACTTTCT	1121
Q	y	902	CAGGCACAGTCAGGCCCATCTGTCGTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC	961
D	b	1122	CAGGCACAGTCAGGCCCATCTGTCGTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC	1181
Q	y	962	CACCTCTGGATCATTTGGATGGGCTTTACGAAGCAGAGATGCGAGGAAGATGCTTGACATAC	1021
D	b	1182	CACCTCTGGATCATTTGGATGGGCTTTACGAAGCAGAGATGCGAGGAAGATGCTTGACATAC	1241
Q	y	1022	TGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGACAGCATGCGTPACC	1081
D	b	1242	TGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGACAGCATGCGTPACC	1301
Q	y	1082	AGGGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGAAGGGGTGTGACACCT	1141
D	b	1302	AGGGGGAAGTCACCGAGAAGATGATGTGTGAGGCATCCCGAAGGGGTGTGACACCT	1361
Q	y	1142	GCCAGGCTGACAGTGTGTGGCCCCCTGATGTACCAATGTACCAGTGGCATGTGTTGGGCA	1201
D	b	1362	GCCAGGCTGACAGTGTGTGGCCCCCTGATGTACCAATGTACCAGTGGCATGTGTTGGGCA	1421
Q	y	1202	TCGTTAGCTGGGCTATGGCTGCGGGGGCCGAGCACCCAGAGATATACCAAGTCT	1261
D	b	1422	TCGTTAGCTGGGCTATGGCTGCGGGGGCCGAGCACCCAGAGATATACCAAGTCT	1481
Q	y	1262	CAGCCTATCTCAACTGGATCTACAATGTCTGGAAGCTGAGCTG	1305
D	b	1482	CAGCCTATCTCAACTGGATCTACAATGTCTGGAAGCTGAGCTG	1525
RESULT 9				
AR232520				
LOCUS AR232520 2079 bp DNA linear PAT 20-DEC-2002				
DEFINITION Sequence 1 from patent US 6455668.				
ACCESSION AR232520				
VERSION AR232520.1 GI:27274657				
KEYWORDS				
SOURCE Unknown.				
ORGANISM Unclassified.				
REFERENCE 1 (bases 1 to 2079)				
AUTHORS Mack,D., Gish,K.C. and Wilson,K.E.				
TITLE Methods of diagnosing colorectal cancer, compositions, and methods of screening for colorectal cancer modulators				
JOURNAL Patent: US 6455668-A 1 24-SEP-2002;				
FEATURES Location/Qualifiers				
source 1..2079				
/organism="unknown"				
/mol_type="genomic DNA"				
 ORIGIN				
Query Match 99.0%; Score 1292; DB 6; Length 2079;				
Best Local Similarity 99.9%; Pred. No. 9.5e-299;				
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;				
Q	y	3	GGATCTGCAGTGATCAACCTCTGAAAGCTTCGATGTCAAAACCCCTGGCAAAACCCCG	62
D	b	216	GGATCTGCAGTGATCAACCTCTGAAAGCTTCGATGTCAAAACCCCTGGCAAAACCCCG	275

63 TATCCCATGGAGACCTTCAGAAAG-GTGGGGATCCCATCATCATAGCACTACTGAGCC 121
DB
276 TATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
QY
122 TGGGAGATATCATATGTTGTTGCTCTCATCAAGGTGATTCGATATAATACTACTTCC 181
DB
336 TGGGAGATATCATATGTTGTTGCTCTCATCAAGGTGATTCGATATAATACTACTTCC 395
QY
182 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGAGCGAGAGCTGGACT 241
DB
396 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGAGCGAGAGCTGGACT 455
QY
242 GTCCCTTGGGGAGGACGAGAGAGCTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 301
DB
456 GTCCCTTGGGGAGGACGAGAGAGCTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 515
QY
302 CAGTCGGCTCTCCAGAGACGATCCACATGCGAGGTCTGAGTCCGCCACAGGGAAT 361
DB
516 CAGTCGGCTCTCCAGAGACGATCCACATGCGAGGTCTGAGTCCGCCACAGGGAAT 575
QY
362 GGTTCCTCTGCTGTTTTCGACAACTTCACAGAGCTTCGCTGAGACAGCTGTAGGAGA 421
DB
576 GGTTCCTCTGCTGTTTTCGACAACTTCACAGAGCTTCGCTGAGACAGCTGTAGGAGA 635
QY
422 TGGGCTACAGAGCAAAACCCATTCAGAGCTGTGAGATGGCCGACAGGATCTGG 481
DB
636 TGGGCTACAGAGCAAAACCCATTCAGAGCTGTGAGATGGCCGACAGGATCTGG 695
QY
482 ATGTTGTTGAAATCACAGAAACAGCAGGAGCTTCGATCGGAATCAAGTGGGCCCT 541
DB
696 ATGTTGTTGAAATCACAGAAACAGCAGGAGCTTCGATCGGAATCAAGTGGGCCCT 755
QY
542 GTCTCTAGGCTCCCTGTTCTCCCTGCACTGTCTGCTGTGGAGAGCTGTAGACCC 601
DB
756 GTCTCTAGGCTCCCTGTTCTCCCTGCACTGTCTGCTGTGGAGAGCTGTAGACCC 815
QY
602 CCGCTGTGTGGTGGGAGAGGAGCTCTGTGATTCCTGGCTTGGCAGGTCAAGCATCC 661
DB
816 CCGCTGTGTGGTGGGAGAGGAGCTCTGTGATTCCTGGCTTGGCAGGTCAAGCATCC 875
QY
662 AGTAGACAAACAGACAGCTCTGTGAGGAGAGCATCTCGAAGAGCTGTGCTCTCAGG 721
DB
876 AGTAGACAAACAGACAGCTCTGTGAGGAGAGCATCTCGAAGAGCTGTGCTCTCAGG 935
QY
722 CAGCCCACTGCTTCAGGAACATACGATGTTGTTCAATGGAAGTGGGCGAGCTCAG 781
DB
936 CAGCCCACTGCTTCAGGAACATACGATGTTGTTCAATGGAAGTGGGCGAGCTCAG 995
QY
782 ACAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATCAACC 841
DB
996 ACAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATCAACC 1055
QY
842 CCAATGTACCCCAAGAACATGATCGCCCTCATGAAGCTGAGTCCCATCTCACTTCT 901
DB
1056 CCAATGTACCCCAAGAACATGATCGCCCTCATGAAGCTGAGTCCCATCTCACTTCT 1115
QY
902 CAGGCACTGAGGCGCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 961
DB
1116 CAGGCACTGAGGCGCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1175
QY
962 CACTGTGATCATTTGATGGGCTTTTACGAACAGAAATGGAGGAGATGCTGACATAC 1021
DB
1176 CACTGTGATCATTTGATGGGCTTTTACGAACAGAAATGGAGGAGATGCTGACATAC 1235
QY
1022 TGTGAGGCGCTCAGTTCAGGTCATTGACAGACACGCTGCAATGACAGAGATGGGTACC 1081
DB
1236 TGTGAGGCGCTCAGTTCAGGTCATTGACAGACACGCTGCAATGACAGAGATGGGTACC 1295
QY
1082 AGGGGAAATCAACCGAAGATGATGTGTGCGAGGATCCCGAAGGGGTGTGGACACT 1141
DB
1296 AGGGGAAATCAACCGAAGATGATGTGTGCGAGGATCCCGAAGGGGTGTGGACACT 1355
QY
1142 GCCAGGGTACAGTGTGGGCGCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGGCA 1201

1356 GCAGGGTACAGTGTGGGCGCCCTGATGTACCAATCTGACCAATGTGGTGGGCA 1415
QY
1202 TCGTTAGCTGGGCGCTATGGCTCGGGGGCCCGAGCACCCAGAGATATACCAAGGTCT 1261
DB
1416 TCGTTAGCTGGGCGCTATGGCTCGGGGGCCCGAGCACCCAGAGATATACCAAGGTCT 1475
QY
1262 CAGCCTATCTCACTGAGATCAATGTCTGAAGGCTGAGCTG 1305
DB
1476 CAGCCTATCTCACTGAGATCAATGTCTGAAGGCTGAGCTG 1519
RESULT 10
AX207967 2079 bp DNA linear PAT 31-AUG-2001
LOCUS
DEFINITION
Sequence 71 from Patent WO0157194.
AX207967
ACCESSION
AX207967.1 GI:15422563
VERSION
KEYWORDS
Homo sapiens (human)
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Madison, E.L., Ong, E.O. and Yeh, J.C.
TITLE
Nucleic acid molecules encoding transmembrane serine proteases, the
encoded proteins and methods based thereon
JOURNAL
Patent: WO 0157194-A 71 09-AUG-2001;
CORVAS INTERNATIONAL, INC. (US)
FEATURES
Location/Qualifiers
1..2079
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
251..1522
/note="unnamed protein product; Nucleotide sequence
encoding transmembrane protease, serine 4 (TPRSS4)"
/codon_start=1
/protein_id="C60389.1"
/db_xref="GI:15422564"
/db_xref="EMBL:CA60389"
/translation="MSNPANPVSPPSPESVGIPIIALSLASIIIVLVKVLVD
KYFLCGQLHPIPRKQCDGLDPLGDEHCVKSPSPGPAVRLSKDSTLQVL
DATGNWFSACDFNFTALAEACQMGYSSTPTFAVEIGDQLDVVEITENSQGL
RVNSGHPCLSGSLVSLHCLACKSLKTPRVVGGEEASVDSPWQVSIQYDKHVGGL
SILDPHWLTAACHFRKHPTDVFNVKVRAGSKLGSPPSLAVAKIIIEFNPMYKDN
IALMKQPLPTFTSGTVPICLPFDEBELTPATLWIGWFTKONGKMSDILLQASV
QVLDSTRCNADDAVQGEVTEKMKACGIPGGVDTCQDSDGGLPMYQSDQWHVVGVSW
GYGCGSPSPGVYTKVSAVLNINYNVWKAEI"
ORIGIN
Query Match 99.0%; Score 1292; DB 6; Length 2079;
Best Local Similarity 99.9%; Pred. No. 9.5e-299;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 3 GGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCACACCCCTCGGAAACCCCG 62
DB 216 GGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCACACCCCTCGGAAACCCCG 275
QY 63 TATCCCCATGAGACCTTCAGAAAG-GTGGGGATCCCATCATCATAGCACTACTGAGCC 121
DB 276 TATCCCCATGAGACCTTCAGAAAGTGTGGGATCCCATCATCATAGCACTACTGAGCC 335
QY 122 TGGGAGATATCATATGTTGTTGCTCTCATCAAGGTGATTCGATATAATACTACTTCC 181
DB 336 TGGGAGATATCATATGTTGTTGCTCTCATCAAGGTGATTCGATATAATACTACTTCC 395
QY 182 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGAGCGAGAGCTGGACT 241
DB 396 TCTGGGCGAGCTCTCCACATTCATCCGAGGAAGCAGCTGTGTGAGCGAGAGCTGGACT 455
QY 242 GTCCCTTGGGGAGGACGAGAGAGCTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 301

456	Db	GTCCCTTTGGGGAGGAGACGAGGAGCATTGTGTCAAGAGCTTCCCGAAGGGGCTTCGAGTGG	515
302	Qy	CAGTCCGCTCTCCAAGAGACCGATCCACACTGCAAGTGTCTGGACTCGGCCACAGGGAAC	361
516	Db	CAGTCCGCTCTCCAAGAGACCGATCCACACTGCAAGTGTCTGGACTCGGCCACAGGGAAC	575
362	Qy	GGTTCTCTGGCTGTTCGACAACTTCCACAGAGCTCTCGCTGACACAGCTGTAGGCAGA	421
576	Db	GGTTCTCTGGCTGTTCGACAACTTCCACAGAGCTCTCGCTGACACAGCTGTAGGCAGA	635
422	Qy	TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGGCCACAGACCATGTGG	481
636	Db	TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGGCCACAGACCATGTGG	695
482	Qy	ATGTGTGTGAATCACAGAAACAGCCAGAGCTTCGCATGCGGAATCAAGTGGGCGCT	541
696	Db	ATGTGTGTGAATCACAGAAACAGCCAGAGCTTCGCATGCGGAATCAAGTGGGCGCT	755
542	Qy	GTCTCTCAGGCTCCCTGGTCTCCCTGCACGTCTTTGGCTGTGGGAAGAGCCTGAAGACCC	601
756	Db	GTCTCTCAGGCTCCCTGGTCTCCCTGCACGTCTTTGGCTGTGGGAAGAGCCTGAAGACCC	815
602	Qy	CCGCTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGTACC	661
816	Db	CCGCTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGTACC	875
662	Qy	AGTACGACAAACAGCACGCTCTGTGGAGGAGCATCCTGGACCCCACTGGAGCTCAG	721
876	Db	AGTACGACAAACAGCACGCTCTGTGGAGGAGCATCCTGGACCCCACTGGAGCTCAG	935
722	Qy	CAGCCACTCTTCAGGAACATACCATGTGTTCAACTGGAGGTTGGGAGGCTCAG	781
936	Db	CAGCCACTCTTCAGGAACATACCATGTGTTCAACTGGAGGTTGGGAGGCTCAG	995
782	Qy	ACAAATGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCAC	841
996	Db	ACAAATGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTCAC	1055
842	Qy	CCATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGCAAGTCCCACTCACTTCT	901
1056	Db	CCATGTACCCCAAGACAAATGACATCGCCCTCATGAGCTGCAAGTCCCACTCACTTCT	1115
902	Qy	CAGGCACAGTCAGGCCCATCTGCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACC	961
1116	Db	CAGGCACAGTCAGGCCCATCTGCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACC	1175
962	Qy	CACCTCTGGATCTTGGATGGGCTTTACGAGCAGAAATGAGGGAAGATGCTGACATAC	1021
1176	Db	CACCTCTGGATCTTGGATGGGCTTTACGAGCAGAAATGAGGGAAGATGCTGACATAC	1235
1022	Qy	TGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGACAGACGATCGTACC	1081
1236	Db	TGCTGCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCAATGACAGACGATCGTACC	1295
1082	Qy	AGGGGAAGTCCCGGAGAATGATGTGCGAGCATCCCGGAGGGGTTGGACACCT	1141
1296	Db	AGGGGAAGTCCCGGAGAATGATGTGCGAGCATCCCGGAGGGGTTGGACACCT	1355
1142	Qy	GCCAGGTTGACAGTGTGGGCCCTGTATGATACCAATCTGACCAAGTGGCATGTGGTGGGCA	1201
1356	Db	GCCAGGTTGACAGTGTGGGCCCTGTATGATACCAATCTGACCAAGTGGCATGTGGTGGGCA	1415
1202	Qy	TCGTTAGCTGGGGTATGGCTCGGGGGCCGAGCACCCAGCAGTATACCAAGTCT	1261
1416	Db	TCGTTAGCTGGGGTATGGCTCGGGGGCCGAGCACCCAGCAGTATACCAAGTCT	1475
1262	Qy	CAGCCTATCTCAACTGGATCTACAAATGTCTGGGAAGGCTGAGCTG	1305
1476	Db	CAGCCTATCTCAACTGGATCTACAAATGTCTGGGAAGGCTGAGCTG	1519

RESULT 11
AF216312

LOCUS	AF216312	2079 bp	mRNA	linear	PRI 07-FEB-2000
DEFINITION	Homo sapiens type II membrane serine protease mRNA, complete cds.				
ACCESSION	AF216312				
VERSION	AF216312.1	GI:6911218			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Sneekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.				
TITLE	Mt-SP2, a novel type II membrane serine protease expressed in trachea, colon, and small intestine: identification, cloning, and chromosomal localization				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2079)				
AUTHORS	Sneekens, S.S., Lorimer, D.D., Wang, E., Hou, J. and Linnevers, C.				
TITLE	Direct Submission				
JOURNAL	Submitted (14-DEC-1999) Axy's Pharmaceuticals, Inc, 180 Kimball Way, South San Francisco, CA 94080, USA				
FEATURES	Location/Qualifiers				
source	1..2079				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
CDS	251..1522				
	/note="Mt-SP2"				
	/codon_start=1				
	/product="type II membrane serine protease"				
	/protein_id="AAF31436.1"				
	/db_xref="GI:6911219"				
	/translation="MSNPQNPVPRPSPESVGIPIIIALLSLASIIIVVLLIKVLD KYFLCGQLPHITPKQLDCGLDEDEHCVKSFPEGPAVAVRLSKDRSTLOVL DSATGNWFACDHTPTEALACTACROMGYSSKPTPRAVEIGPDQDLDVVEITENSOGL RWNSSGCLSGSLHGLACGSKLTKPRVVGGEASVDSMPWQVSIQVDKQHVGG SULDHPWLTAACHKFKHPTDVENKVRAGSKLGSFSLAVAKILLIENPMYPKND IALMLKQLPFTSGIVRPICLPFDEBELIPATPLWIIWGFTKQNGKMSDILHQAQSV QVIDSTRNADDAIQGEVTEKMMCGIPEGGVDITCQDSSGGLPMYQSDQWVVVGIYSV GYCGCGPSPTPGVTVKVSAYLNIYVNWKAEL"				
ORIGIN					
Query Match	99.0%; Score 1292; DB 9; Length 2079;				
Best Local Similarity	99.9%; Pred. No. 9, 5e-299;				
Matches 1303; Conservative	0; Mismatches 0; Indels 1; Gaps 1;				
QY	3	GGATCTCTGACAGTGATCAACCTCTGAAAGCCTTCGATGTCAAACCCCTGGCAAAACCCCG	62		
Db	216	GGATCTCTGACAGTGATCAACCTCTGAAAGCCTTCGATGTCAAACCCCTGGCAAAACCCCG	275		
QY	63	TATCCCATGGAGACCTTCAGAAAG-GTGGGATCCCATCATCATAGCACTACTGAGCC	121		
Db	276	TATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGCC	335		
QY	122	TGGCGAGTATCATCATTTGGTTGTCTCTCATCAAGGTGATCTGGATAAATACTACTTCC	181		
Db	336	TGGCGAGTATCATCATTTGGTTGTCTCTCATCAAGGTGATCTGGATAAATACTACTTCC	395		
QY	182	TCGTGGGGGAGCCTTCACACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTCGACT	241		
Db	396	TCGTGGGGGAGCCTTCACACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTCGACT	455		
QY	242	GTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAGGCCCTGCAAGTGG	301		
Db	456	GTCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAGGCCCTGCAAGTGG	515		
QY	302	CAGTCGCGCTCTCCAGGAGCCGATCCACACTCAGCTGTGTGACCTGGCCACAGGAACT	361		
Db	516	CAGTCGCGCTCTCCAGGAGCCGATCCACACTCAGCTGTGTGACCTGGCCACAGGAACT	575		
QY	362	GGTTCTCTGCCCTGTTCGAACTTCACAGAAGCTTCGCTGAGACAGCTGTGAGCAGA	421		
Db	576	GGTTCTCTGCCCTGTTCGAACTTCACAGAAGCTTCGCTGAGACAGCTGTGAGCAGA	635		

QY 422 TGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTGG 481
DB 536 TGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTGG 695
QY 482 ATGTGTGTAATACAGAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCTT 541
DB 696 ATGTGTGTAATACAGAAACAGCAGGAGCTTCGATCGGAACTCAAGTGGGCTT 755
QY 542 GTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTCTGTGGAAAGAGCTCAAGACCC 601
DB 755 GTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTCTGTGGAAAGAGCTCAAGACCC 815
QY 602 CCCGTGTGGTGGGAGAGAGCCCTCTGTGGATCTTGGCCCTTGGCAGGTGAGCATCC 661
DB 816 CCCGTGTGGTGGGAGAGAGCCCTCTGTGGATCTTGGCCCTTGGCAGGTGAGCATCC 875
QY 662 AGTAGGCAAAACAGCAGCTCTGTGGAGGAGCATCCCTGGACCCCACTGGGTCTCAAG 721
DB 876 AGTAGGCAAAACAGCAGCTCTGTGGAGGAGCATCCCTGGACCCCACTGGGTCTCAAG 935
QY 722 CAGCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAG 781
DB 936 CAGCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGAGGCTCAG 995
QY 782 ACAAACTGGGAGCTTCCATCCCTGGCTGTGGGCAAGATCATCATATTGAATTCAAAC 841
DB 996 ACAAACTGGGAGCTTCCATCCCTGGCTGTGGGCAAGATCATCATATTGAATTCAAAC 1055
QY 842 CCATGTACCCCAAGAAACATACGATCGCCCTCATGAAGCTGCAGTCCCACTCACTTCT 901
DB 1056 CCATGTACCCCAAGAAACATACGATCGCCCTCATGAAGCTGCAGTCCCACTCACTTCT 1115
QY 902 CAGGACAGTCAGGCGCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 961
DB 1116 CAGGACAGTCAGGCGCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACC 1175
QY 962 CACTCTGATCATTTGATGGGCTTTACGAAGCAGATGGAAGGAGATGCTGACATAC 1021
DB 1176 CACTCTGATCATTTGATGGGCTTTACGAAGCAGATGGAAGGAGATGCTGACATAC 1235
QY 1022 TGCTGCAGGCTCAGTCAGGCTCATTCAGACACACCGTGAATGACAGATGCGTACC 1081
DB 1236 TGCTGCAGGCTCAGTCAGGCTCATTCAGACACACCGTGAATGACAGATGCGTACC 1295
QY 1082 AGGGGGAATCACCGAGAGATGATGTGTCAGGATCCCGGAGAGGGGTGTGACACCT 1141
DB 1296 AGGGGGAATCACCGAGAGATGATGTGTCAGGATCCCGGAGAGGGGTGTGACACCT 1355
QY 1142 GCCAGGCTGACAGTGTGGGCCCCCTGTATGTATACCAATCTGACAGTGTGGTGGCA 1201
DB 1356 GCCAGGCTGACAGTGTGGGCCCCCTGTATGTATACCAATCTGACAGTGTGGTGGCA 1415
QY 1202 TCGTTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTCT 1261
DB 1416 TCGTTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTCT 1475
QY 1262 CAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1305
DB 1476 CAGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1519

RESULT 12
LOCUS AX092380
DEFINITION Sequence 111 from Patent WO0116318.
ACCESSION AX092380
VERSION AX092380.1 GI:13444504
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1

AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.B., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 111 08-MAR-2001;
Genentech, Inc. (US)
FEATURES Location/Qualifiers
source 1..2063
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 96.8%; Score 1263; DB 6; Length 2063;
Best Local Similarity 98.8%; Pred. NO. 8.4e-292;
Matches 1288; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 62
DB 218 GGATCCTGACAGTATCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 277
QY 63 TATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGAGCT 122
DB 278 TATCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGAGCT 337
QY 123 GCGAGTATCATCATGTGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCT 182
DB 338 GCGAGTATCATCATGTGTGTCTCATCAAGGTGATCTGGATAAATACTACTTCT 397
QY 183 CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAATG 242
DB 398 CTGCGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGAATG 457
QY 243 TCCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGACTTCCCGAGGAGCTGAGTGGC 302
DB 458 TCCCTTGGGGAGGAGCAGGAGCACTGTGTCAAGACTTCCCGAGGAGCTGAGTGGC 517
QY 303 AGTCGCGCTCTCCAAAGGACCGATCCACACTGACAGCTGTGTGACTCGGCCACAGGGAATG 362
DB 518 AGTCGCGCTCTCCAAAGGACCGATCCACACTGACAGCTGTGTGACTCGGCCACAGGGAATG 577
QY 363 GTTCTGTGCTTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAT 422
DB 578 GTTCTGTGCTTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAT 637
QY 423 GGGCTACAGCAGCAAAACCCACTTCAGAGCTGTGAGATTTGGCCACAGACAGGATCTGGA 482
DB 638 GGGCTACAG-----CAGAGCTGTGAGATTTGGCCACAGACAGGATCTGGA 682
QY 483 TGTGTGTAATCACAGAAACAGCCAGAGCTTCGCAATGCGGAATCTCAAGTGGGCCCTG 542
DB 683 TGTGTGTAATCACAGAAACAGCCAGAGCTTCGCAATGCGGAATCTCAAGTGGGCCCTG 742
QY 543 TCTCTCAGGCTCCCTGCTCCCTGCACTGTCTGCTGTGGAAAGAGCTTGAAGACCCC 602
DB 743 TCTCTCAGGCTCCCTGCTCCCTGCACTGTCTGCTGTGGAAAGAGCTTGAAGACCCC 802
QY 603 CCGTGTGTGGTGGGAGGAGGCGCTCTGTGGAATTTTGGCCCTTGGCAGTCAAGATCA 662
DB 803 CCGTGTGTGGTGGGAGGAGGCGCTCTGTGGAATTTTGGCCCTTGGCAGTCAAGATCA 862
QY 663 GTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCTCAGCGC 722
DB 863 GTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTGGACCCCACTGGTCTCTCAGCGC 922
QY 723 AGCCCACTGCTTCAGGAAACATACCGATGTTCAACTGGAAGGTGCGGCGAGGCTCAGA 782
DB 923 AGCCCACTGCTTCAGGAAACATACCGATGTTCAACTGGAAGGTGCGGCGAGGCTCAGA 982
QY 783 CAACTGGGAGCTTCCCATCCCTGGCTGTGGCAGAGATCATCATGATTCATCAACC 842
DB 983 CAACTGGGAGCTTCCCATCCCTGGCTGTGGCAGAGATCATCATGATTCATCAACC 1042

Qy	843	CATGTACCCCAAAGACAATCACATPGCCCTCATGAAGCTCAGTTCCTCACTCACTTCTC	902
Db	1043	CATGTACCCCAAAGACAATGACATCGCCCTCATGAAGCTCAGTTCCTCACTCACTTCTC	1102
Qy	903	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACCCC	962
Db	1103	AGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGAGCTCACTCCAGCCACCCC	1162
Qy	963	ACTCTGGATCATTTGGATGGGGCTTTTACGAAGACAGAAATGGAGGAAGATGCTGACATACT	1022
Db	1163	ACTCTGGATCATTTGGATGGGGCTTTTACGAAGACAGAAATGGAGGAAGATGCTGACATACT	1222
Qy	1023	GCTGCAGGCGTCAGTCCAGTCAATTGACAGCACACGCTGCAATGCAGACGATCGGTACCA	1082
Db	1223	GCTGCAGGCGTCAGTCCAGTCAATTGACAGCACACGCTGCAATGCAGACGATCGGTACCA	1282
Qy	1083	GGGGGAAGTCACCGAAGAATGATGTGTGCAGGCAATCCCGAAGGGGGTGTGCACACCTG	1142
Db	1283	GGGGGAAGTCACCGAAGAATGATGTGTGCAGGCAATCCCGAAGGGGGTGTGCACACCTG	1342
Qy	1143	CNAGGTGACAGTGTGTGGGCCCTTGATGTACCAATCTGACATGGCATGTGTGGGCAT	1202
Db	1343	CNAGGTGACAGTGTGTGGGCCCTTGATGTACCAATCTGACATGGCATGTGTGGGCAT	1402
Qy	1203	CGTTAGCTGGGGCTATGGCTGCGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTC	1262
Db	1403	CGTTAGCTGGGGCTATGGCTGCGGGGCCGAGCACCCAGGAGTATACACCAAGGTCTC	1462
Qy	1263	AGCCTATCTCAACTGGATCTPACAATGTCTCGAAGGCTGAGCTG	1305
Db	1463	AGCCTATCTCAACTGGATCTPACAATGTCTCGAAGGCTGAGCTG	1505
RESULT 13			
AX376262			
LOCUS	AX376262	2063 bp	DNA linear PAT 01-MAR-2002
DEFINITION	Sequence 329 from Patent WO0168848.		
ACCESSION	AX376262		
VERSION	AX376262.1 GI:19170522		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 Baker, K. P., Chen, J., Desnoyers, L., Goddard, A., Godowski, P. J., Gurney, A. L., Pan, J., Smith, V., Watanabe, C. K., Wood, W. I. and Zhang, Z.		
AUTHORS	Secreted and transmembrane polypeptides and nucleic acids encoding the same		
TITLE	Patent: WO 0168848-A 329 20-SEP-2001; Genentech, Inc. (US)		
JOURNAL	Location/Qualifiers		
FEATURES	1..2063		
source	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match 96.8%; Score 1263; DB 6; Length 2063;			
Best Local Similarity 98.8%; Pred. No. 8.4e-292;			
Matches 1288; Conservative 0; Mismatches 0; Indels 15; Gaps 1;			
Qy	3	GGATCTCGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGGCAAAACCCCG	62
Db	218	GGATCTCGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGGCAAAACCCCG	277
Qy	63	TATCCCCATGGACCTTCAGAAGCTGGGGATCCCCCATCATATAGCACTACTGAGCCT	122
Db	278	TATCCCCATGGAGACCTTCAGAAGCTGGGGATCCCCCATCATATAGCACTACTGAGCCT	337
Qy	123	GGCAGTATCATCATTTGGTTCTCTCTCATCAAGGTGATTTCTGGATAAATACTTCTCT	182

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 96.8%; Score 1263; DB 6; Length 2063;
Best Local Similarity 98.8%; Pred. No. 8.4e-292;
Matches 1288; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 3 GGATCCTGACAGTATCAACTCTGAAACAGCCTCGATGTCGACCAACCCCG 62
Db 218 GGATCCTGACAGTATCAACTCTGAAACAGCCTCGATGTCGACCAACCCCG 277
QY 63 TATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATATAGCACTACTGAGCCT 122
Db 278 TATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATATAGCACTACTGAGCCT 337
QY 123 GCGAGTATCATCTGTTGGTGTCTCATCAAGGTGATCTCGATAAATACTTCTCT 182
Db 338 GCGAGTATCATCTGTTGGTGTCTCATCAAGGTGATCTCGATAAATACTTCTCT 397
QY 193 CTGGGGAGCCTCTCCATCTATCCCGAGGAGAGCTGTGTGACGGAGAGCTGGACTG 242
Db 398 CTGGGGAGCCTCTCCATCTATCCCGAGGAGAGCTGTGTGACGGAGAGCTGGACTG 457
QY 243 TCCCTTGGGGAGGACGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 302
Db 458 TCCCTTGGGGAGGACGAGGAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 517
QY 303 AGTCGCGCTCTCCAAGGACCGATCCACACTGACAGGTGTGGACTCGGCCACAGGGAACTG 362
Db 518 AGTCGCGCTCTCCAAGGACCGATCCACACTGACAGGTGTGGACTCGGCCACAGGGAACTG 577
QY 363 GTTCTCTGCTGTTTCGACAACTTCACAGAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 422
Db 578 GTTCTCTGCTGTTTCGACAACTTCACAGAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 637
QY 423 GGGGTACAGAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTGGA 482
Db 638 GGGGTACAG-----CAGAGCTGTGGAGATTGGCCACAGCAGGATCTGGA 682
QY 483 TGTGTTGAAATCAGAAAAACAGCAGGAGCTTCGATGCGGAACTCAAGTGGGCGCTG 542
Db 683 TGTGTTGAAATCAGAAAAACAGCAGGAGCTTCGATGCGGAACTCAAGTGGGCGCTG 742
QY 543 TCTCTCAGGCTCCCTGTGCTCCCTGCACTGTCTTGTGCTGTGGGAGAGCCTGAAGACCCC 602
Db 743 TCTCTCAGGCTCCCTGTGCTCCCTGCACTGTCTTGTGCTGTGGGAGAGCCTGAAGACCCC 802
QY 603 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCA 662
Db 803 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTTCAGCATCCA 862
QY 663 GTACGACAAACAGACGCTCTGTGAGGAGAGCTCTGAGACCCCTGCTGGCTCTCAGGCG 722
Db 863 GTACGACAAACAGACGCTCTGTGAGGAGAGCTCTGAGACCCCTGCTGGCTCTCAGGCG 922
QY 723 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAGA 782
Db 923 AGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAGA 982
QY 783 CAATGCGGAGCTTCCCATCTCGGCTGTGGCCCAAGATCATCATATTGAATTCACACC 842
Db 983 CAATGCGGAGCTTCCCATCTCGGCTGTGGCCCAAGATCATCATATTGAATTCACACC 1042
QY 843 CATGTACCCCAAGACATGATCGCCCTCATGAGCTGAGTCCCACTCAGCTTCTC 902
Db 1043 CATGTACCCCAAGACATGATCGCCCTCATGAGCTGAGTCCCACTCAGCTTCTC 1102
QY 903 AGGCACAGTCAAGGCCATCTCTCTGCTGCTCTTTTGAAGGAGCTCACTCAGGCCACCC 962
Db 1103 AGGCACAGTCAAGGCCATCTCTCTGCTGCTCTTTTGAAGGAGCTCACTCAGGCCACCC 1162

Search completed: June 8, 2004, 01:15:54

Job time : 5333 secs

QY 963 ACTCTGGATCATTTGGATGGGGCTTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACT 1022
Db 1163 ACTCTGGATCATTTGGATGGGGCTTTTACGAAGCAGAAATGGAGGGAAGATGTCTGACATACT 1222
QY 1023 GCTGCAGGGGTCAAGTCCAGGTCAATTGACAGCACACCGGTGCAATGACAGATGCGGTACCA 1082
Db 1223 GCTGCAGGGGTCAAGTCCAGGTCAATTGACAGCACACCGGTGCAATGACAGATGCGGTACCA 1282
QY 1083 GGGGGAAGTACCCGAGGAAGATGATGTGCAGGCATCCCGGAAGGGGGTGTGGACACCTG 1142
Db 1283 GGGGGAAGTACCCGAGGAAGATGATGTGCAGGCATCCCGGAAGGGGGTGTGGACACCTG 1342
QY 1143 CCAGGGTGCAGTGTGGGGCCCTGATGTACCAATCTGTACCAGTGGCATGTGTGGGCAT 1202
Db 1343 CCAGGGTGCAGTGTGGGGCCCTGATGTACCAATCTGTACCAGTGGCATGTGTGGGCAT 1402
QY 1203 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTC 1262
Db 1403 CGTTAGCTGGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACACCAAGGTCTC 1462
QY 1263 AGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1305
Db 1463 AGCCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1505

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47; Search time 23 Seconds
(without alignments)
976.404 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLKRP.....VYTKVSAYLNWIYVWKAEI 435

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID
1	2338	99.8	435	US-09-008-271A-6
2	2337	99.8	437	US-08-851-588-8
3	2196	93.8	423	US-08-656-002-2
4	2188	93.4	406	US-08-851-588-6
5	677.5	28.9	492	US-09-685-166A-895
6	676.5	28.9	492	US-09-342-749-2
7	676.5	28.9	492	US-09-691-840-2
8	655.5	28.0	454	US-09-518-046-2
9	598.5	25.1	417	US-09-820-002-4
10	584	24.9	455	US-08-261-416-2
11	580	24.8	376	US-09-820-002-2
12	574	24.5	416	US-09-000-846-2
13	571	24.4	798	1 US-08-200-900A-2
14	571	24.4	798	5 PCT-US94-00616-2
15	568	24.3	418	US-08-508-448C-25
16	568	24.3	418	US-09-370-838-82
17	568	24.3	418	US-09-370-838-83
18	566	24.2	418	US-09-370-838-62
19	558.5	23.8	283	3 US-08-807-151-1
20	558.5	23.8	283	4 US-09-478-957-1
21	531.5	22.7	232	1 US-08-508-448C-19
22	512.5	21.9	256	2 US-09-027-337-3
23	512.5	21.9	256	4 US-08-644-600-3
24	512.5	21.9	256	4 US-08-654-600A-3
25	511	21.8	638	2 US-08-681-151-3
26	507.5	21.7	255	3 US-08-944-483-67
27	503.5	21.5	285	4 US-09-023-942A-26

28	486	20.8	248	3	US-08-944-483-63	Sequence 63, Appl
29	483	20.6	314	4	US-09-023-942A-6	Sequence 6, Appl
30	482.5	20.6	284	4	US-09-387-375-7	Sequence 7, Appl
31	478	20.4	314	3	US-09-008-271A-3	Sequence 3, Appl
32	478	20.4	314	4	US-09-907-794A-257	Sequence 257, App
33	478	20.4	314	4	US-09-905-125A-257	Sequence 257, App
34	478	20.4	314	4	US-09-902-775A-257	Sequence 257, App
35	477	20.4	312	4	US-09-023-942A-4	Sequence 4, Appl
36	475	20.3	407	4	US-09-734-675-4	Sequence 4, Appl
37	472.5	20.2	317	4	US-09-386-629-7	Sequence 7, Appl
38	472.5	20.2	317	4	US-09-907-794A-263	Sequence 263, App
39	472.5	20.2	317	4	US-09-905-125A-263	Sequence 263, App
40	472.5	20.2	317	4	US-09-902-775A-263	Sequence 263, App
41	470	20.1	902	4	US-09-644-600-10	Sequence 10, Appl
42	470	20.1	902	4	US-09-654-600A-10	Sequence 10, Appl
43	467.5	20.0	290	4	US-09-386-653A-7	Sequence 7, Appl
44	466.5	19.9	405	4	US-09-734-675-2	Sequence 2, Appl
45	465.5	19.9	250	3	US-08-944-483-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1
US-09-008-271A-6
; Sequence 6, Application US/09008271A
; Patent No. 6203979
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/09/008,271A
; FILING DATE: 16-Jan-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLN0113
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-008-271A-6

Query Match 99.8%; Score 2338; DB 3; Length 435;
Best Local Similarity 99.8%; Pred. No. 1.2e-234; Mismatches 0; Indels 0; Gaps 0;
Matches 434; Conservative 1;

QY 1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 60
DB 1 MDPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 60

QY 61 LCGQPLHPIPRKQLCDGELDCPLGEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 120
DB 61 LCGQPLHPIPRKQLCDGELDCPLGEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 120

QY 121 WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGP 180
DB 121 WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 240
DB 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 240

QY 241 AAHCFRKHDTVFNWVKRAGSKLGSFPLSAVAKIIIEFNPMYKDNDAIAMLKLOFPLTF 300
DB 241 AAHCFRKHDTVFNWVKRAGSKLGSFPLSAVAKIIIEFNPMYKDNDAIAMLKLOFPLTF 300

QY 301 SGTVPICLPFFDEELTPATPLMIIGMFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVPICLPFFDEELTPATPLMIIGMFTKQNGKMSDILLQASVQVIDSTRCNADDAY 360

QY 361 QGEVTEKMCAGIPGGVDTCQDSDGGLMYQSDQHVHVGIVSGYCGGPGSTFGVTKV 420
DB 361 QGEVTEKMCAGIPGGVDTCQDSDGGLMYQSDQHVHVGIVSGYCGGPGSTFGVTKV 420

QY 421 SAYLNWIYVWKAEL 435
DB 421 SAYLNWIYVWKAEL 435

RESULT 2
US-09-851-588-8
; Sequence 8, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-8

Query Match 99.8%; Score 2337; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.5e-234; Mismatches 0; Indels 0; Gaps 0;
Matches 434; Conservative 0;

QY 2 DPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 61
DB 4 DPDSQPLNSLDVPLKPRIPMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFL 63

QY 62 CGOPLHFIIPRKQLCDGELDCPLGEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 121

DB 64 CGOPLHFIIPRKQLCDGELDCPLGEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGN 123
QY 122 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGPC 181
DB 124 FSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGPC 183
QY 182 LGSLSVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 241
DB 184 LGSLSVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTA 243

QY 242 AHCFRKHDTVFNWVKRAGSKLGSFPLSAVAKIIIEFNPMYKDNDAIAMLKLOFPLTF 301
DB 244 AHCFRKHDTVFNWVKRAGSKLGSFPLSAVAKIIIEFNPMYKDNDAIAMLKLOFPLTF 303

QY 302 GTVRPICLPFFDEELTPATPLMIIGMFTKQNGKMSDILLQASVQVIDSTRCNADDAY 361
DB 304 GTVRPICLPFFDEELTPATPLMIIGMFTKQNGKMSDILLQASVQVIDSTRCNADDAY 363

QY 362 GEVTEKMCAGIPGGVDTCQDSDGGLMYQSDQHVHVGIVSGYCGGPGSTFGVTKV 421
DB 364 GEVTEKMCAGIPGGVDTCQDSDGGLMYQSDQHVHVGIVSGYCGGPGSTFGVTKV 423

QY 422 AYLNIWIYVWKAEL 435
DB 424 AYLNIWIYVWKAEL 437

RESULT 3
US-09-656-002-2
; Sequence 2, Application US/09656002
; Patent No. 6455668
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt
; APPLICANT: Wilson, Keith
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-69108/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/656,002
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/525,993
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: US 09/493,444
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/US 00/07044
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 93.8%; Score 2196; DB 4; Length 423;
Best Local Similarity 97.1%; Pred. No. 6.9e-220;
Matches 408; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 16 PLRKPRIIPMETFRKVGIPITIIALLSLASIIIVVLLIKVILDKYFLCGOPLHFIIPRKQLC 75
DB 4 PCANPVSPPRPSESUGIPITIIALLSLASIIIVVLLIKVILDKYFLCGOPLHFIIPRKQLC 63

QY 76 DGELDCPLGEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGNWFSACFDNFTALAE 135
DB 64 DGELDCPLGEDEHCHVKSPPEGPAVAVRLSKORSTLOVLDSATGNWFSACFDNFTALAE 123

QY 136 TACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGPCLSGSLVSLHCLAG 195
DB 124 TACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELMRNSSGPCLSGSLVSLHCLAG 183

QY 196 KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTAACFRKHDTVFNW 255
DB 184 KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLTAACFRKHDTVFNW 243

QY 256 VRAGSDKLGSPFSLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEE 315
Db 244 VRAGSDKLGSPFSLAVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEE 303
QY 316 LTPATPLMIIGWGTQKQNGKSDILLOASVOVIDSTRCNADDAAYQGEVTEKMWCAIPE 375
Db 304 LTPATPLMIIGWGTQKQNGKSDILLOASVOVIDSTRCNADDAAYQGEVTEKMWCAIPE 363
QY 376 GGVDTCCGSDGGLMYQSDQWHVVGIVSGWYCGCGPSTPGVYTKVSAYLWYNNWKAE 435
Db 364 GGVDTCCGSDGGLMYQSDQWHVVGIVSGWYCGCGPSTPGVYTKVSAYLWYNNWKAE 423

RESULT 4

US-09-851-588-6
; Sequence 6, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David C.
; APPLICANT: Gish, Kurt E.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; FILE REFERENCE: A-68829-1/DJB/JJD/RMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-851-588-6

Query Match 93.4%; Score 2188; DB 4; Length 406;
Best Local Similarity 100.0%; Pred. No. 4.4e-219;
Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 30 VGPIIIALLSLASIIIVVLIKVIDKYFLCGQPLHFIPRKQLCDGELDCPLGEDBEH 89
Db 1 VGPIIIALLSLASIIIVVLIKVIDKYFLCGQPLHFIPRKQLCDGELDCPLGEDBEH 60
QY 90 CVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAEATACROMGYSSKPTF 149
Db 61 CVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAEATACROMGYSSKPTF 120
QY 150 RAVEIGPQDLDVVEITENSQELMRNSSGFCLSGLVSLHCLACGKSLKTPRVVGGEEA 209
Db 121 RAVEIGPQDLDVVEITENSQELMRNSSGFCLSGLVSLHCLACGKSLKTPRVVGGEEA 180
QY 210 SVDSWPMQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFVFNKVRAGSDKLGSPFSL 269
Db 181 SVDSWPMQVSIQYDKQHVCGGSIIDPHWVLTAAHCFRKHDTVFVFNKVRAGSDKLGSPFSL 240
QY 270 AVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATPLMIIGWGF 329
Db 241 AVAKIIIEFNPMYKNDIALMKLQPLTFSGTVRPICLPFFDEELTPATPLMIIGWGF 300
QY 330 TKQNGKGVSDILLOASVOVIDSTRCNADDAAYQGEVTEKMWCAIPEGGVDTCCGSDGGL 389
Db 301 TKQNGKGVSDILLOASVOVIDSTRCNADDAAYQGEVTEKMWCAIPEGGVDTCCGSDGGL 360
QY 390 MYQSDQWHVVGIVSGWYCGCGPSTPGVYTKVSAYLWYNNWKAE 435
Db 361 MYQSDQWHVVGIVSGWYCGCGPSTPGVYTKVSAYLWYNNWKAE 406

RESULT 5

US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqul
; APPLICANT: Hendersot, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895
Query Match 28.9%; Score 677.5; DB 4; Length 492;
Best Local Similarity 39.4%; Pred. No. 1e-61;
Matches 151; Conservative 56; Mismatches 127; Indels 49; Gaps 13;
QY 75 CDGELDCPLGEDBEHCHVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAE 134
Db 133 CDGELDCPLGEDBEHCHVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTALAE 134
QY 135 ETACROMGYSSKPTFRAVEIGPQDLDVVEITENSQELMRNSSG-----PCL 182
Db 182 RAACRDMGY--KXNFYSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLSVSLHCLACGKSL---KTPRVVCGEASVDSDWPMQVSIQYDKQHVCGGSIIDPHWVLT 239
Db 233 SKAVSLRCLACGVNLNSRQSRIVGGSALPCAWPQVSLHVQNVHVCVCGSIITPEWIV 292
QY 240 TAAHCFRKH--TDVFNKVRAGSDKLGSPF---PSLAVAKIIIEFNPMY---PKNDIAL 291
Db 293 TAAHCFRKH--TDVFNKVRAGSDKLGSPF---PSLAVAKIIIEFNPMY---PKNDIAL 348
QY 292 MKLQPLTFSGTVRPICLPFFDEELTPATPLMIIGWGTQKQNGKSDILLOASVOVIDS 351
Db 349 MKLQPLTFSGTVRPICLPFFDEELTPATPLMIIGWGTQKQNGKSDILLOASVOVIDS 407
QY 352 TRCNADDAAYQGEVTEKMWCAIPEGGVDTCCGSDGGLMYQSDQ--WHVVGIVSGWYCGCG 410
Db 408 QRCNSRYVDNLITPAMI CAGFLQGNVDSQCGSDGGLMYQSDQ--WHVVGIVSGWYCGCG 467
QY 411 PSTPGVYTKVSAYLWYNNWKAE 433
Db 468 AYRFGVYGVNWFPTDIYRQMK 490

RESULT 6

US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.

```

; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match      28.9%; Score 676.5; DB 3; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.3e-61;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTTEALA 134
Db 133 CDGVSHCPGGEDENRCVRLY--GP-----NFIQVYSSQRKSNHPVQCDDWENYNG 181
QY 135 ETACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRNNSG-----PCL 182
Db 182 RAACRDMGY--KNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLSLVSLHCLACGKSL---KTPRVVGGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVL 239
Db 233 SKAVVSLRACIACGVNLNRSRQSRIVGGESALPGAWPMQVSLHVQNVHVCGGSIITPEWIV 292
QY 240 TAAHCFERKH-TDVFNNKVRAGSDKLGSP----PSLAVAKIIIIIEFNPMY----PKNDIAL 291
Db 293 TAAHCFERKPLNPNWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNDIAL 348
QY 292 MKLQPLFTSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDS 351
Db 349 MKLQPLFTNDLVKPVCLPNPGMWLQPEQLCWIISGWGATEEK-GKTSEVLNAAKVLIIET 407
QY 352 TRCNADDAVQGEVTEKMCAGIPEGVDTCQDGGPLMYQSDQ-WHVVGVISVWGYGCGG 410
Db 408 QRCNSRYVYDNLITPAMICAGFLQGVNDSQDGGPLVTSKNNIWWLIGDTSWGSCKAK 467
QY 411 PSTPGVYTKVSAYLNIWYNWKA 434
Db 468 AYRPGVYGVNVMVFTDIYRQMEAD 491

RESULT 7
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 644419
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match      28.0%; Score 655.5; DB 3; Length 454;
Best Local Similarity 35.7%; Pred. No. 1.7e-59;
Matches 158; Conservative 73; Mismatches 149; Indels 63; Gaps 18;

QY 22 IPMETFRKVGIPFIALLSLAIIIVVLIKVILD---KY-----YFLCQPHFIPRQOL 74
Db 44 LPPEVFSQSSSLGIALL-----ILALIGLHFDCSGKYRCRSFKC---IELITR--- 92
QY 75 CDGELDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTTEALA 134
Db 93 CDGVSDCKGGEDEYRC-----VRVGQNAVQLVFTAA--SWKTMCSDDMKGHYA 139
QY 135 ETACROMGYSS-----KPTFRAVEIGPDQDLVDVVEITENSQELMRNNSG 182
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TADG-12
; OTHER INFORMATION: protein
US-09-518-046-2

Query Match      28.9%; Score 676.5; DB 3; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.3e-61;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTTEALA 134
Db 133 CDGVSHCPGGEDENRCVRLY--GP-----NFIQVYSSQRKSNHPVQCDDWENYNG 181
QY 135 ETACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRNNSG-----PCL 182
Db 182 RAACRDMGY--KNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLSLVSLHCLACGKSL---KTPRVVGGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVL 239
Db 233 SKAVVSLRACIACGVNLNRSRQSRIVGGESALPGAWPMQVSLHVQNVHVCGGSIITPEWIV 292
QY 240 TAAHCFERKH-TDVFNNKVRAGSDKLGSP----PSLAVAKIIIIIEFNPMY----PKNDIAL 291
Db 293 TAAHCFERKPLNPNWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNDIAL 348
QY 292 MKLQPLFTSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDS 351
Db 349 MKLQPLFTNDLVKPVCLPNPGMWLQPEQLCWIISGWGATEEK-GKTSEVLNAAKVLIIET 407
QY 352 TRCNADDAVQGEVTEKMCAGIPEGVDTCQDGGPLMYQSDQ-WHVVGVISVWGYGCGG 410
Db 408 QRCNSRYVYDNLITPAMICAGFLQGVNDSQDGGPLVTSKNNIWWLIGDTSWGSCKAK 467
QY 411 PSTPGVYTKVSAYLNIWYNWKA 434
Db 468 AYRPGVYGVNVMVFTDIYRQMEAD 491

RESULT 7
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 644419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2

Query Match      28.9%; Score 676.5; DB 3; Length 492;
Best Local Similarity 39.1%; Pred. No. 1.3e-61;
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCVKSFPEGPAPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTTEALA 134
Db 133 CDGVSHCPGGEDENRCVRLY--GP-----NFIQVYSSQRKSNHPVQCDDWENYNG 181
QY 135 ETACROMGYSSKPTFRAVEIGPDQDLVDVVEITENSQELMRNNSG-----PCL 182
Db 182 RAACRDMGY--KNFYSSQ-----GIVDDSGSTSFMKLNTSAGNVDIYKLYHSDACS 232
QY 183 SGLSLVSLHCLACGKSL---KTPRVVGGEEASVDSWPMQVSIQYDKQHVCGGSIIDPHWVL 239
Db 233 SKAVVSLRACIACGVNLNRSRQSRIVGGESALPGAWPMQVSLHVQNVHVCGGSIITPEWIV 292
QY 240 TAAHCFERKH-TDVFNNKVRAGSDKLGSP----PSLAVAKIIIIIEFNPMY----PKNDIAL 291
Db 293 TAAHCFERKPLNPNWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNDIAL 348
QY 292 MKLQPLFTSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDS 351
Db 349 MKLQPLFTNDLVKPVCLPNPGMWLQPEQLCWIISGWGATEEK-GKTSEVLNAAKVLIIET 407
QY 352 TRCNADDAVQGEVTEKMCAGIPEGVDTCQDGGPLMYQSDQ-WHVVGVISVWGYGCGG 410
Db 408 QRCNSRYVYDNLITPAMICAGFLQGVNDSQDGGPLVTSKNNIWWLIGDTSWGSCKAK 467
QY 411 PSTPGVYTKVSAYLNIWYNWKA 434
Db 468 AYRPGVYGVNVMVFTDIYRQMEAD 491

RESULT 7
US-09-691-840-2
; Sequence 2, Application US/09691840
; Patent No. 644419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-749-2
```

Db 140 NVACQALGFPPSVSDNLRVSSLEQGFREFFVSDHLLPDDKVTALHHSVYVREG---CA 196
Qy 183 SGLSVSLHCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 241
Db 197 SGHVVTLOCTACGHRGYSRRIVGGNMSLLSQWPAQSLQFGYHLGCGSVITPLWITA 256
Qy 242 AHCFRKHDTVF---NWKVRAGSDKL--GSFPLAVAKIIIEFNPMY-PK--DNDIALMK 293
Db 257 AHCV---YDLYLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
Qy 294 LQFPLTFSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVOVIDSTR 353
Db 311 LAGPLTFNEMIQVCLPNSSENFPGKVCWTSWGAT-EDGGDASPVLNHAAPLISNKI 369
Qy 354 CNADDAVQGEVTERKMCAGIPGGVDTCCGSGGGLMYQSDQ-WHVGVIVSWGYG 408
Db 370 CNHRDVGGLIISPSMLCAGYLTGGVDSQCGSGGGLVQERRLWKLVGATSGIGCAEVN 429
Qy 413 TPGVTVKVSAYLNWLYNWKAE 435
Db 430 KPGVTVRTVTSFLDWIHEQMERDL 452

RESULT 9
US-09-820-002-4
; Sequence 4, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weinlu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: US 2001/0194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 417
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-4

Query Match 25.1%; Score 588.5; DB 4; Length 417;
Best Local Similarity 31.5%; Pred. No. 1.4e-52;
Matches 140; Conservative 60; Mismatches 161; Indels 83; Gaps 11;
Qy 19 KPRIMETFRKVGIPITIIALLSLASIIIVVVLTKVILD---KY---YFLCGQLFPIRKLCDGE 78
Db 15 RPKVAALT---AGTLLLTGAASWAIVALLR----- 45
Qy 79 LDCPLGEDEEHCVKSPFPGPAVALRSLKDRSTLQVLDATGNWFSACFNFTEALTAETAC 138
Db 46 -----SDQE-----PLYPGVSSADARLWFDKTEGTRLLCSRSRNVAVGLSC 90
Qy 139 ROMGVSSKPTFRPRAVEIGPDQDLVDVEITEN-----SQELRMENSSGGLS 183
Db 91 EEMGF-----LEAL---THSELVDVRTAGANGTSFGFCVDSGRPLPHTQRLLEVISFSDCPR 142
Qy 184 GSLVSLHCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 242
Db 143 GRFLAAICQDCGRRLKPLVDRIYGGRTSLGRWPMQVSLRYDGAHLGCGSLLSGDWLTA 202
Qy 243 HCFRKHDTVF--WKVRAGSDKLGSFPLAVAKIIIEFNPMY-----KNDIALMK 293
Db 203 HCFPERNVLRSWRVFAVAQAQSPHGLQGVQAVVHGGYLPFADPNSENSNDIALVH 262
Qy 294 LQFPLTFSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVOVIDSTR 353
Db 263 LSSPLPLEYIQVCLPAAQALVDGKICTVTGWGNT-QYVGQAQVGLQEARVPIISNDV 321

Qy 354 CNADDAVQGEVTERKMCAGIPGGVDTCCGDSGGGLMYQ-----SDQWHVGVIVSWGYG 408
Db 322 CNGADFYGNQIKPMFCAGYPEGIDACQDSGGPFVCEDSISRTPRWRLCGIVSWGTGC 381
Qy 409 GGPSTPGVTVKVSAYLNWLYNWK 432
Db 382 ALAQKPGVTVKVSDFREWFQAIK 405

RESULT 10
US-09-261-416-2
; Sequence 2, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: C'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
US-09-261-416-2

Query Match 24.9%; Score 584; DB 3; Length 455;
Best Local Similarity 34.2%; Pred. No. 4.8e-52;
Matches 153; Conservative 73; Mismatches 150; Indels 72; Gaps 21;
Qy 22 IPMETFRKVGIPITIIALLSLASIIIVVVLTKVILD---KY---YFLCGQLFPIRKL 74
Db 44 LPFEVFSQSSSLGIIAL-----ILALAIGLGIHFDCGKYRCRSPKC---IBLITR--- 92
Qy 75 CDGELDCPLGEDEEHCVKSPFPGPAVALRSLKDRSTLQVLDATGNWFSACFNFTEALA 134
Db 93 CDGVSDCKDGEDEVRC-----VRVGGQNAVQLQVFTAA--SWKTCSDDWKGYHA 139
Qy 135 ETACROMGYSS-----KPTFRAVEIGPDQDLVDVEITENSQELRMENSSGGL 182
Db 140 NVACQALGFPPSVSDNLRVSSLEQGFREFFVSDHLLPDDKVTALHHSVYVREG---CA 196
Qy 183 SGLSVSLHCLACG-KSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSILDPHWLTA 241
Db 197 SGHVVTLOCTACGHRGYSRRIVGGNMSLLSQWPAQSLQFGYHLGCGSVITPLWITA 256
Qy 242 AHCFRKHDTVF---NWKVRAGSDKL--GSFPLAVAKIIIEFNPMY-PK--DNDIALMK 293
Db 257 AHCV---YDLYLPKSWTIQVGLVSLDNPAPSHLVEKIV---YHSKYKPKRLGNDIALMK 310
Qy 294 LQFPLTFSGTVRPICLPFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVOVIDSTR 353
Db 311 LAGPLTFNEMIQVCLPNSSENFPGKVCWTSWGAT-EDGGDASPVLNHAAPLIS--- 366
Qy 354 CNADDAVQGEVTERKMCAGIPGGVDTCCGSGGGLMYQSDQ-WHVGVIVSWGYG 407
Db 367 -NKDLPQGRVRWHHLPLHALRGLPDGWRNWSQCGSGGLVQERRLWKLVGATSGIG 425
Qy 408 CCGPSTPGVTVKVSAYLNWLYNWKAE 435
Db 426 CADVKNFPGVTVRTVTSFLDWIHEQMERDL 453

RESULT 11
US-09-820-002-2
; Sequence 2, Application US/09820002

```

; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 376
; ORGANISM: HUMAN
; ORGANISM: HUMAN
; US-09-820-002-2

Query Match          24.8%; Score 580; DB 4; Length 376;
Best Local Similarity 30.8%; Pred. No. 9.5e-52;
Matches 132; Conservative 57; Mismatches 146; Indels 94; Gaps 9;

QY 19 KPRIPETKRGVPIIIALLSLAIIIVVLKIVLDKYFLCGQLHFIKRLQCDGE 78
DB 15 RKVVAALT---AGTLLLTGAASWAIIVALLR-----45

QY 79 LDCPLGEDEHCKVSPPEGPAVAVRLSKDRSTLOVLSATGNWFSACFDNFTALAEATAC 138
DB 46 -----SDQE-----PLYPQVSSADARLMVFDKTEGTRWLLCSSRSNARVAGLSC 90

QY 139 ROMGYSSKTFRAVEIGPDQDLVDVEITENSQELMRNSSGCLSGSLVSLHCLACG-KS 197
DB 91 EENGFLS-----DCPRGRFLAACQDCGRK 116

QY 198 LKTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWLTAACFRKHTDVEN-WKV 256
DB 117 LPVDRIVGGRDTSLGSRWPQVSLRYDGAHLGSLLSGDWVLTAAHCFPERNRVLSRWV 176

QY 257 RAGSDKLGSPFSLAVAKIIIEFNMPY-----KNDIALMKLPFLFSGTVRIC 308
DB 177 FAGAVASPHGLQLGQVAVVHGGVLPFRDPSNSENNDIALVHLSPLPTEYIQVVC 236

QY 309 LPFFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDDAYQGVETKM 368
DB 237 LPAAGQALVDGKICTVTGNGNT-QYVGGQAGVLQEARVPIISNDVNGADFYGNQIKPKM 295

QY 369 MCAGIEGGVDTCCGDSGGPLMYQ-----SDQHVHVGIVSWGYGCGGSTPGVTKVSAY 423
DB 296 FCAGYEGGIDACQDSGGPFVCEDSISRTPRWRLCGIVSWGTGALAAKPGVYTKVSDF 355

QY 424 LNWIVNWK 432
DB 356 REWIFQAIK 364

RESULT 12
US-09-000-846-2
; Sequence 2, Application US/09000846
; Patent No. 5981830
; GENERAL INFORMATION:
; APPLICANT: WU, QINGYU
; APPLICANT: SADLER, JASPER
; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
; TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US

```

```

; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,846
; FILING DATE: 30-DEC-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/866,058
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: LEROVITZ, RICHARD M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: BERLX 65P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 416 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-000-846-2

Query Match          24.5%; Score 574; DB 2; Length 416;
Best Local Similarity 30.8%; Pred. No. 4.6e-51;
Matches 135; Conservative 62; Mismatches 151; Indels 90; Gaps 10;

QY 30 VGIPIIALLSLAIIIVVLKIVLDKYFLCGQLHFIKRLQCDGELDCPLGDEEH 89
DB 22 VGTLLFTGICASWAIIVLLO-----SDQE- 48

QY 90 CVKSPPEGPAVAVRLSKDRSTLOVLSATGNWFSACFDNFTALAEATACRQNGY----- 143
DB 49 -----PLYQVQLSPGDSRLAVFDKTEGTRWLLCSSRSNARVAGLCEEMGFLALAH 100

QY 144 -----SSKPTFAVEIG-----PDQDLVDVEITENSQELMRNSSGCLSGSLVSL 189
DB 101 SELDVRTAGANGTSGFPCVDEGGLRLAQLLDVISVD-----CPRGRFLTA 147

QY 190 HCLACG-KSKLTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWLTAACFRKH 248
DB 148 TCQDCGRKRLPVDRIVGQDSLSGRWPQVSLRYDGTDLHLCGSLSGDWLTAACFCPE 207

QY 249 TDYFN-WKVPSAGSKLGSFSLAVAKIIIEFNMPY-----KNDIALMKLPFLT 299
DB 208 NRVLRSRVRVAGAVARTSPHVLQVQVAVIYHGGVLPFRDPTIDENSNDIALVHLSPL 267

QY 300 FSGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDA 359
DB 268 LTEYIQVCLPAAQALVDGKICTVTGNGNT-QYVGGQAGVLQEARVPIISNEVCNSPDF 326

QY 360 YQGEVTEKMCAGIEGGVDTCCGDSGGPLMYQ-----SDQHVHVGIVSWGYGCGGSTP 414
DB 327 YGNQIKPKMFCAGYEGGIDACQDSGGPFVCEDSISGTSRWRLCGIVSWGTGALARKP 386

QY 415 GVTYKVSAYLNWIVNWK 432
DB 387 GVTYKVTDFREWIFKAIK 404

RESULT 13
US-08-200-300A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38

```

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02140
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/200,900A
 ; FILING DATE: 23-FEB-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meinerdt, Maureen C.
 ; REGISTRATION NUMBER: 31,544
 ; REFERENCE/DOCKET NUMBER: GI 5201-FWC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170 X8574
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 798 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US94-00616-2

Query Match 24.4%; Score 571; DB 1; Length 798;
 Best Local Similarity 35.0%; Pred. No. 2.6e-50;
 Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;
 ;
 ; 69 IPRKQLCDGELDCPLGEDHECHVKSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDN 128
 ; 435 IPLVNLCDGFPCHCKDGSDEAHCVRLLF-NGTDSGLVQFR--IQSI-----WHVACAEN 485
 ;
 ; 129 FTEALAEATACRQMGY-----SSKPTFRAVEIGPDQDLQVVE-----ITENSQELMRNMSGP 180
 ; 486 WTTQISDDVCCQLGLGTGNSVPTF-STGGGPVYNLTAPNGSLILTPSQO----- 535
 ;
 ; 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSWPQVSIQYDKQHVCGGSILD 234
 ; 536 CLEDSLILLQCNYSKCGKLVQEVSPKIVGSDSREGANPWVVALYFDDQVCGASLVS 595
 ; 235 PHWLTAHC-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIIIEFNPMY----PKDND 288
 ; 596 RDWLVAACHVYGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRRKND 655
 ; 289 IALMKLOPPLTSGTVRPTCLPFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQV 348
 ; 656 IAMHLEKVNITYDIQIPLPEENQVFPGRICSIAGWALIQ-GSTADVLQEAADVPL 714
 ; 349 IDSTRNADDAQGEVTEKMCAGIPEGVDTCQDGGGLMYQ-SDQWHVGVISWGYG 407
 ; 715 LSNEKQ-QOMPEYNITENVCAGYAGVDSCQDGGGLMCOENNRWLLAGVTSFGVQ 773
 ; 408 CGGPSTPGVYTKVSAYLNWI 427
 ; 774 CALPNREGVYARVPRFTEWI 793

US-08-200-900A-2

Query Match 24.4%; Score 571; DB 1; Length 798;
 Best Local Similarity 35.0%; Pred. No. 2.6e-50;
 Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;
 ;
 ; 69 IPRKQLCDGELDCPLGEDHECHVKSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDN 128
 ; 435 IPLVNLCDGFPCHCKDGSDEAHCVRLLF-NGTDSGLVQFR--IQSI-----WHVACAEN 485
 ;
 ; 129 FTEALAEATACRQMGY-----SSKPTFRAVEIGPDQDLQVVE-----ITENSQELMRNMSGP 180
 ; 486 WTTQISDDVCCQLGLGTGNSVPTF-STGGGPVYNLTAPNGSLILTPSQO----- 535
 ;
 ; 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSWPQVSIQYDKQHVCGGSILD 234
 ; 536 CLEDSLILLQCNYSKCGKLVQEVSPKIVGSDSREGANPWVVALYFDDQVCGASLVS 595
 ; 235 PHWLTAHC-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIIIEFNPMY----PKDND 288
 ; 596 RDWLVAACHVYGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRRKND 655
 ; 289 IALMKLOPPLTSGTVRPTCLPFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQV 348
 ; 656 IAMHLEKVNITYDIQIPLPEENQVFPGRICSIAGWALIQ-GSTADVLQEAADVPL 714
 ; 349 IDSTRNADDAQGEVTEKMCAGIPEGVDTCQDGGGLMYQ-SDQWHVGVISWGYG 407
 ; 715 LSNEKQ-QOMPEYNITENVCAGYAGVDSCQDGGGLMCOENNRWLLAGVTSFGVQ 773
 ; 408 CGGPSTPGVYTKVSAYLNWI 427
 ; 774 CALPNREGVYARVPRFTEWI 793

RESULT 14

PCT-US94-00616-2
 ; Sequence 2, Application PC/TUS9400616
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
 ; NUMBER OF SEQUENCES: 33
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/00616
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 798 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US94-00616-2
 ;
 ; Query Match 24.4%; Score 571; DB 5; Length 798;
 ; Best Local Similarity 35.0%; Pred. No. 2.6e-50;
 ; Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;
 ;
 ; 69 IPRKQLCDGELDCPLGEDHECHVKSPPEGPAVAVRLSKDRSTLQVLD SATGNWFSACFDN 128
 ; 435 IPLVNLCDGFPCHCKDGSDEAHCVRLLF-NGTDSGLVQFR--IQSI-----WHVACAEN 485
 ;
 ; 129 FTEALAEATACRQMGY-----SSKPTFRAVEIGPDQDLQVVE-----ITENSQELMRNMSGP 180
 ; 486 WTTQISDDVCCQLGLGTGNSVPTF-STGGGPVYNLTAPNGSLILTPSQO----- 535
 ;
 ; 181 CLSGSLVSLHC--LACGKSLKT-----PRVVGGEASVDSWPQVSIQYDKQHVCGGSILD 234
 ; 536 CLEDSLILLQCNYSKCGKLVQEVSPKIVGSDSREGANPWVVALYFDDQVCGASLVS 595
 ; 235 PHWLTAHC-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIIIEFNPMY----PKDND 288
 ; 596 RDWLVAACHVYGRNMEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRRKND 655
 ; 289 IALMKLOPPLTSGTVRPTCLPFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQV 348
 ; 656 IAMHLEKVNITYDIQIPLPEENQVFPGRICSIAGWALIQ-GSTADVLQEAADVPL 714
 ; 349 IDSTRNADDAQGEVTEKMCAGIPEGVDTCQDGGGLMYQ-SDQWHVGVISWGYG 407
 ; 715 LSNEKQ-QOMPEYNITENVCAGYAGVDSCQDGGGLMCOENNRWLLAGVTSFGVQ 773
 ; 408 CGGPSTPGVYTKVSAYLNWI 427
 ; 774 CALPNREGVYARVPRFTEWI 793
 ;
 ; RESULT 15
 ; US-08-508-448C-25
 ; Sequence 25, Application US/08508448C
 ; Patent No. 5804410
 ; GENERAL INFORMATION:
 ; APPLICANT: Kazuyoshi YAMAOKA et al.
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
 ; TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wenderoth, Lind & Ponack
 ; STREET: 805 Fifteenth Street, N.W., #700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/508,448C

Search completed: June 7, 2004, 08:42:39
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 20:10:49 ; Search time 573 Seconds
(without alignments)
9675.216 Million cell updates/sec

Title: US-10-030-688-1
Perfect score: 1305
Sequence: 1 atggtatctgacagtgatca.....atgtctggaagctgagctg 1305

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1305	100.0	1305	4	AAD02556	Aad02556 Human ser
2	1303.4	99.9	2038	2	AXX87154	Axx87154 Human pro
3	1303	99.8	1314	7	ABX76354	Abx76354 Lung canc
4	1303	99.8	1314	9	ADB80524	Adb80524 Ovarian c
5	1303	99.8	2081	6	AAI72976	Aai72976 CJA8 pref
6	1303	99.8	2307	6	ABS76529	Abs76529 cDNA enco
7	1303	99.8	2307	7	ACF12937	Acf12937 Human cer
8	1301.8	99.8	2121	6	AAI64284	Aai64284 Human ser
9	1298.2	99.5	2137	4	ADL13114	Adl13114 Human mem
10	1297.4	99.4	1479	4	AAD02557	Aad02557 Human gen
11	1294	99.2	2165	6	ABZ35523	Abz35523 Human ser
12	1292	99.0	2070	3	AAZ90471	Aaz90471 Cancer sp
13	1292	99.0	2079	4	AAI13169	Aai13169 Human tra
14	1292	99.0	2079	6	AAI72975	Aai72975 CJA8 cDNA
15	1292	99.0	2079	7	ABS57763	Abs57763 cDNA enco
16	1292	99.0	2079	8	ACF35041	Acf35041 Human tra
17	1292	99.0	2079	9	ADE48026	Ade48026 DNA enco
18	1263	96.8	2063	3	AAZ37099	Aaz37099 Human PRO
19	1263	96.8	2063	4	AAF54396	Aaf54396 DNA enco
20	1263	96.8	2063	4	AAZ46089	Aaz46089 Human DNA
21	1263	96.8	2063	4	AAF92113	Aaf92113 Human PRO
22	1263	96.8	2063	6	ABS74433	Abs74433 Human PRO
23	1263	96.8	2063	6	ABK11090	Abk11090 cDNA enco

ALIGNMENTS

RESULT 1	
AAD02556	
ID	AAD02556 standard; cDNA; 1305 BP.
XX	AC
XX	AAD02556;
XX	02-MAY-2001 (first entry)
XX	DE
XX	Human seripancrin cDNA.
XX	Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
KW	arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
KW	osteoporosis; aberrant wound healing; angiogenesis; diabetes;
KW	inflammatory disorder; stroke; cardiovascular disease; gene therapy;
KW	vaccine; cytostatic; cerebroprotective; vulnary; osteopathic; ss.
XX	OS
XX	Homo sapiens.
XX	OS
XX	Key
XX	Location/Qualifiers
XX	1..1305
FT	/*tag= a
FT	/product= "Human seripancrin protein"
FT	/note= "The coding region does not include stop codon"
FT	/partial
XX	WO200104141-A2.
XX	XX
XX	18-JAN-2001.
XX	XX
PP	04-JUL-2000; 2000WO-EP006211.
XX	XX
PR	12-JUL-1999; 99EP-00113428.
XX	XX
XX	(MERE) MERCK PATENT GMBH.
XX	XX
PI	Suendermann B, Hofmann U, Matzku S, Wilbert O;
PI	WPI; 2001-147177/15.
DR	P-PSDB; AAY72558.
XX	XX
PPT	New extracellular serine protease Seripancrin, useful for treating
PPT	cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
PPT	diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound
PPT	healing.
XX	XX
PS	Claim 5; Page 37-39; 45pp; English

Abx78692 Human PRO
Aca75664 Novel hum
Aca71144 Human sec
Acc87672 Human sec
Acc87058 Human sec
Acc04231 Human sec
Aca69562 cDNA enco
Aca90407 Novel hum
Acc89514 Human sec
Aca98305 Novel hum
Aca93947 Human sec
Acd15340 Human sec
Acd08927 Human sec
Acc96847 Human sec
Acf15568 Human sec
Aca72935 Human PRO
Acd03107 Novel hum
Acd01922 Novel hum
Aca92114 Novel hum
Aca89539 cDNA enco
Aca73549 Human sec
Aca05864 Human sec

XX The present invention relates to seripancrin polynucleotides, and
CC polypeptides encoded by them. Seripancrin are members of serine protease
CC family. This protein contains a transmembrane domain, a low density
CC lipoprotein (LDL) domain, protease domain and a scavenger receptor
CC cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the
CC specificity of seripancrin's intra and intermolecular interactions. The
CC polynucleotides and polypeptides of the invention are useful for treating
CC and diagnosing diseases such as arthritis, chronic obstructive pulmonary
CC disorder (COPD), cancer, osteoporosis, aberrant wound healing,
CC angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular
CC diseases. Seripancrin genes are useful in chromosome localisation
CC studies as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists and
CC antagonists useful for treating conditions associated with seripancrin
CC imbalance. These polypeptides are also useful as vaccines. The present
CC sequence is a cDNA coding for seripancrin protein. The seripancrin gene
CC is located on human chromosome 11q22-q23
XX
SQ Sequence 1305 BP; 293 A; 368 C; 372 G; 272 T; 0 U; 0 Other;

Query Match 100.0%; Score 1305; DB 4; Length 1305;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGATCTCTGACAGTGATCAACCTCTGAACAGCCTCTGATGTCACACCCCTCGCAAAACC 60
DB 1 ATGGATCTCTGACAGTGATCAACCTCTGAACAGCCTCTGATGTCACACCCCTCGCAAAACC 60
QY 61 CGATATCCCATGAGACCTTCAGAAAGTGGGGATCCCATCATATAGCACTACTGAGC 120
DB 61 CGATATCCCATGAGACCTTCAGAAAGTGGGGATCCCATCATATAGCACTACTGAGC 120
QY 121 CTGGCGAGTATCATATCTGTGTCTCTCAAGAGTGATCTCGATAAATACTACTTC 180
DB 121 CTGGCGAGTATCATATCTGTGTCTCTCAAGAGTGATCTCGATAAATACTACTTC 180
QY 181 CTCTGGGAGCCTCTCCACTTCATCCCGAGAGAGCTGTGTGACGGAGAGCTGGAC 240
DB 181 CTCTGGGAGCCTCTCCACTTCATCCCGAGAGAGCTGTGTGACGGAGAGCTGGAC 240
QY 241 TGTTCCTCTGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 300
DB 241 TGTTCCTCTGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCGAGTG 300
QY 301 GCAGTCCGCTCTCCAAAGAGCCGATCCACACTGCAAGTGTCTGCACTCGCCACAGGGAAC 360
DB 301 GCAGTCCGCTCTCCAAAGAGCCGATCCACACTGCAAGTGTCTGCACTCGCCACAGGGAAC 360
QY 361 TGTTCCTCTGCTGTTTCAGAACTTCAGAAAGCTCTGCTGAGACAGCCTGTAGGCAG 420
DB 361 TGTTCCTCTGCTGTTTCAGAACTTCAGAAAGCTCTGCTGAGACAGCCTGTAGGCAG 420
QY 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 480
DB 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 480
QY 481 GATTTGTTGAAATCAAGAAACAGCCAGGAGCTTCGATCGGAATCAAGTGGGCC 540
DB 481 GATTTGTTGAAATCAAGAAACAGCCAGGAGCTTCGATCGGAATCAAGTGGGCC 540
QY 541 TGTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTGCTGGGAAGAGCCTGAAGACC 600
DB 541 TGTCTCTCAGGCTCCCTGTCTCCCTGCACTGTCTTGTGCTGGGAAGAGCCTGAAGACC 600
QY 601 CCCGCTGTGGTGGGAGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGATC 660
DB 601 CCCGCTGTGGTGGGAGAGGAGGCTCTGTGGATTCTTGGCCTTGGCAGGTCAAGATC 660
QY 661 CAGTACGAAACAGCAGCCTGTGTGAGGAGGAGCATCTCGACCCCTCGGTCTCTCAG 720
DB 661 CAGTACGAAACAGCAGCCTGTGTGAGGAGGAGCATCTCGACCCCTCGGTCTCTCAG 720

QY 721 GCAGCCCACTGCTTCCAGAAACATACCCGATGTGTTCAACTGGAAGTGGGGCAGGCTCA 780
DB 721 GCAGCCCACTGCTTCCAGAAACATACCCGATGTGTTCAACTGGAAGTGGGGCAGGCTCA 780
QY 781 GACAAACTGGGAGCTTCCATCCCTGGCTGGTGGCCAGATCATCATATTGAATTCAAC 840
DB 781 GACAAACTGGGAGCTTCCATCCCTGGCTGGTGGCCAGATCATCATATTGAATTCAAC 840
QY 841 CCCATGTACCCCAAAACAGATGACATCGGCCCTCATGAAGCTCAGTCCCACTCACTTTC 900
DB 841 CCCATGTACCCCAAAACAGATGACATCGGCCCTCATGAAGCTCAGTCCCACTCACTTTC 900
QY 901 TCAGGCACAGTTCAGGCCCATCTGTCTGCTCTTCTTGTATGAGAGCTCACTCCAGCCACC 960
DB 901 TCAGGCACAGTTCAGGCCCATCTGTCTGCTCTTCTTGTATGAGAGCTCACTCCAGCCACC 960
QY 961 CCACTCTGGATCAATTGGATGGGCTTTACGAAGCAATGGAGGAGATGTGTACATA 1020
DB 961 CCACTCTGGATCAATTGGATGGGCTTTACGAAGCAATGGAGGAGATGTGTACATA 1020
QY 1021 CTGCTGAGGCTCAGTCCAGGTCATTGACAGCACACCGTGCATTCAGACATGCGTAC 1080
DB 1021 CTGCTGAGGCTCAGTCCAGGTCATTGACAGCACACCGTGCATTCAGACATGCGTAC 1080
QY 1081 CAGGGGAGTTCACCCAGAGATGATGTGTGAGGATCCCGAGGGGGTGTGGACACC 1140
DB 1081 CAGGGGAGTTCACCCAGAGATGATGTGTGAGGATCCCGAGGGGGTGTGGACACC 1140
QY 1141 TGCCAGGGTGACAGTGGTGGGCCCCCTGATGTATCAATCTGACAGTGGCATGTGGTGGGC 1200
DB 1141 TGCCAGGGTGACAGTGGTGGGCCCCCTGATGTATCAATCTGACAGTGGCATGTGGTGGGC 1200
QY 1201 ATCCTTAGCTGGGCTATGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTTC 1260
DB 1201 ATCCTTAGCTGGGCTATGCTCGGGGGCCCGAGCACCCAGGAGTATACACCAAGTTC 1260
QY 1261 TCAGCCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
DB 1261 TCAGCCCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305

RESULT 2

AX87154
ID AX87154 standard; cDNA; 2038 BP.
XX AX87154;
AC AX87154;
XX 27-SEP-1999 (first entry)
DT Human protease HUPW-6 cDNA.
XX Serine protease; human; HUPW-6; cell proliferation; cancer;
KW immune disorder; inflammation; therapy; ss.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH CDS 200..1507
FT /*tag= a
XX WO9936550-A2.
FN 22-JUL-1999.
XX 12-JAN-1999; 99WO-US000655.
XX 16-JAN-1998; 98US-00008271.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Hillman JL, Yue H, Guegler KJ, Corley NC, Tang YT;
PI Shah P;
XX

DR WPI; 1999-430616/36.
XX P-PSDB; AAY06437..
XX Novel human protease molecules useful in the treatment of developmental
PT disorders and/or cancers.
XX Claim 8; Page 86-87; 90pp; English.
XX This nucleotide sequence codes for HUPM-6 (see AAY06437), a novel human
CC protease. HUPM-6 cDNA was initially identified in Incyte Clone 1337018
CC from the colon cDNA library COLNN0713 using a computer search for amino
CC acid sequence alignments. The present sequence is a consensus sequence
CC derived from overlapping and/or extended nucleic acid consensus sequence
CC Clones 1271725 (TBS7T02), 1337018, 586982 and 588598 (UTRNN001). A
CC fragment comprising nucleotides 900-949 of the present sequence can be
CC used for hybridisation. This sequence encompasses an active site residue.
CC Northern analysis shows expression of HUPM-6 in gastrointestinal, and male
CC an female reproductive cDNA libraries. Approximately 65% of these
CC libraries are associated with neoplastic disorders, and 22% with the
CC immune response. The invention provides 12 new human proteases, i.e. HUPM
CC -1 to -12 (see AAY06432-43), and the polynucleotides encoding them (see
CC AAX87149-60). Also provided are vectors, host cells and methods for
CC producing HUPM polypeptides, as well as agonists and antagonists of
CC Methods for treating or preventing cell proliferative disorders and
CC immune disorders using HUPM or HUPM antagonists are claimed
XX
SQ Sequence 2038 BP; 462 A; 591 C; 569 G; 416 T; 0 U; 0 Other;
Query Match 99.9%; Score 1303.4; DB 2; Length 2038;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGATCTCGACAGTGAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCC 60
DB 200 ATGGATCTCGACAGTGAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCC 259
QY 61 CGTATCCCATCGACAGCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 260 CGTATCCCATCGACAGCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC 319
QY 121 CTGGGAGATCATCATTTGTTGTTCTCTCAAGAGTGAATCTGGATAAATACTACTTC 180
DB 320 CTGGGAGATCATCATTTGTTGTTCTCTCAAGAGTGAATCTGGATAAATACTACTTC 379
QY 181 CTTCTGGGAGCTCTCCACTTCATCCCGAGAGAGAGCTGTGTGACGGAGCTGGAC 240
DB 380 CTTCTGGGAGCTCTCCACTTCATCCCGAGAGAGAGCTGTGTGACGGAGCTGGAC 439
QY 241 TGTCCCTTTGGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTG 300
DB 440 TGTCCCTTTGGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTG 499
QY 301 GCAGTCCGCTCTCCAGAGACGATCCACACTGAGAGTGTGGACTCGGCCACAGGGAAC 360
DB 500 GCAGTCCGCTCTCCAGAGACGATCCACACTGAGAGTGTGGACTCGGCCACAGGGAAC 559
QY 361 TGGTCTCTGCTGCTTTCCAGAACTTCAGAAAGCTTCGCTGAGACAGCTGTAGGAGCAG 420
DB 560 TGGTCTCTGCTGCTTTCCAGAACTTCAGAAAGCTTCGCTGAGACAGCTGTAGGAGCAG 619
QY 421 ATGGGCTACAGCAGCAAAACCCCTTTCAGAGCTGTGGAGATTTGGCCAGACAGGATCTG 480
DB 620 ATGGGCTACAGCAGCAAAACCCCTTTCAGAGCTGTGGAGATTTGGCCAGACAGGATCTG 679
QY 481 GATGTGTGTGAATCACAGAAACAGCAGGAGCTTCGATCGCGAATCTCAAGTGGGCC 540
DB 680 GATGTGTGTGAATCACAGAAACAGCAGGAGCTTCGATCGCGAATCTCAAGTGGGCC 739
QY 541 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAAAGAGCTGAAGACC 600
DB 740 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAAAGAGCTGAAGACC 799
QY 601 CCCCCTGTGGTGGGGAGAGAGCCCTCTGTGGATTTCTTGGCCTTGGCAGGTACGATC 660

DB 800 CCCCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTTCTTGGCTTGGCAGGTACGATC 859
QY 661 CAGTACGACAAACAGCAGCTCTGTGGAGGAGGAGATCTGTGAGCCCACTGGTCTCAGC 720
DB 860 CAGTACGACAAACAGCAGCTCTGTGGAGGAGGAGATCTGTGAGCCCACTGGTCTCAGC 919
QY 721 GCAGCCCACTGCTTTCAGGAAACATACCGATGTGTTCAACTTGAAGGTGCGGGCAGGCTCA 780
DB 920 GCAGCCCACTGCTTTCAGGAAACATACCGATGTGTTCAACTTGAAGGTGCGGGCAGGCTCA 979
QY 781 GACAACTGGGAGCTTCCCATCTCCCTGTGGCCAAAGATCATCATCATGATTTCAAC 840
DB 980 GACAACTGGGAGCTTCCCATCTCCCTGTGGCCAAAGATCATCATCATGATTTCAAC 1039
QY 841 CCCATGTACCCCAAGACAATGATCGCTCATGAAGTGTGAGTTCCTCCACTCCTTTC 900
DB 1040 CCCATGTACCCCAAGACAATGATCGCTCATGAAGTGTGAGTTCCTCCACTCCTTTC 1099
QY 901 TCAGGCACAGTCAAGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCATTCAGGCCACC 960
DB 1100 TCAGGCACAGTCAAGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCATTCAGGCCACC 1159
QY 961 CCCTCTGGATCATTTGGATGGGCTTTACGAGCAGAGATGGAGGAGATGTCTGACATA 1020
DB 1160 CCCTCTGGATCATTTGGATGGGCTTTACGAGCAGAGATGGAGGAGATGTCTGACATA 1219
QY 1021 CTGCTGAGGCGTCAAGTCCAGGTCATTTGACAGCACACGGTGCATTCGACAGCATGCGTAC 1080
DB 1220 CTGCTGAGGCGTCAAGTCCAGGTCATTTGACAGCACACGGTGCATTCGACAGCATGCGTAC 1279
QY 1081 CAGGGGAGTCAAGGAGAGATGTGTGAGGAGATCCCGAAGGGGTGTGGACACC 1140
DB 1280 CAGGGGAGTCAAGGAGAGATGTGTGAGGAGATCCCGAAGGGGTGTGGACACC 1339
QY 1141 TGCAGGCTGACAGTGTGGGCTTCTGATGTACCAATCTGACAGTGGCATGTGGTGGGC 1200
DB 1340 TGCAGGCTGACAGTGTGGGCTTCTGATGTACCAATCTGACAGTGGCATGTGGTGGGC 1399
QY 1201 ATCTTATGCTGGGCTATGGCTGGGGGCTCCGAGACACCCAGGAGTATACCAAGGTC 1260
DB 1400 ATCTTATGCTGGGCTATGGCTGGGGGCTCCGAGACACCCAGGAGTATACCAAGGTC 1459
QY 1261 TCAGCCTATCTCAACTGATCTCAATGTCTGGAAGGCTGAGCTG 1305
DB 1460 TCAGCCTATCTCAACTGATCTCAATGTCTGGAAGGCTGAGCTG 1504
RESULT 3
ABX76354
ID ABX76354 standard; DNA; 1314 BP.
XX AC ABX76354;
XX DT 02-APR-2003 (first entry)
XX DE Lung cancer-associated polynucleotide #218.
XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
XX KW antiinflammatory; antiasthmatic; non-small cell lung cancer; alectasis;
XX KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX OS Unidentified.
XX FN WO200286443-A2.
XX PD 31-OCT-2002.
XX PF 18-APR-2002; 2002WO-US012476.
XX PR 18-APR-2001; 2001US-0284770P.

AAI72976 standard; cDNA; 2081 BP.

AAI72976;

21-AUG-2002 (first entry)

CJA8 preferred cDNA.

Gene; colorectal cancer; CGA7; CJA8; modulating protein; screening;
drug candidate; vaccine; ss.

Homo sapiens.

Key Location/Qualifiers

CDS 215..1528

/*tag= a

/product= "CJA8"

US2002042067-A1.

11-APR-2002.

08-MAY-2001; 2001US-00851588.

17-AUG-2000; 2000US-00642252.

06-SEP-2000; 2000US-00656002.

(MACK/) MACK D.

(GISH/) GISH K C.

(WILSON/) WILSON K E.

Mack D, Gish KC, Wilson KE;

WPI; 2002-453647/48.

P-PSDB; AAG79359.

Screening drug candidates for treating colorectal cancer, comprises
determining the effect of the candidate on the expression profile gene of
CGA7 or CJA8.

Claim 69; Fig 7; 40pp; English.

The sequences given in AAI72973-76 encode the colorectal cancer proteins,
CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal cancer
modulating proteins and have been mapped to chromosomes 2 (CGA7) and 11
(CJA8). These sequences may be used in the method of the invention for
screening drug candidates. The method comprises adding a drug candidate
to a cell that expresses an expression profile gene encoding CGA7, CJA8
or fragments and determining the effect of the drug candidate on the
expression of the expression profile gene. The new methods are used to
screen bioactive agents for the ability to bind to or modulate the
activity of CGA7 or CJA8 and evaluate the effect of a candidate
colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit the
activity of CGA7 or CJA8, respectively, and is used to screen for an
agent that can interfere with the binding of CGA7 or CJA8 to the
antibody. The antibody can be used to treat colorectal cancer. The
antibody or a fragment of it is used to localize a therapeutic group to a
colorectal cancer tissue, where the therapeutic group is a cytotoxic
agent or a radioisotope. Antisense molecules are used to inhibit
colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or CJA8
are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it are used
to elicit an immune response. CGA7 or CJA8 is used to determine the
prognosis of an individual with colorectal cancer. Nucleic acid encoding
CGA7 or CJA8 can be used in vaccines

Sequence 2081 BP; 484 A; 597 C; 576 G; 424 T; 0 U; 0 Other;

Query Match 99.8%; Score 1303; DB 6; Length 2081;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3 GGATCTGACAGTATCAACCTCTGACACGCTCGATGTCACACCCCTCGCAACCCCTCGCAACCCCG 62

Db 223 GGATCTGACAGTATCAACCTCTGAAACAGCCTCGAATGTAAACCCCTCGGCAAAACCCCG 282

QY 63 TATCCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCCATCATCATAGCACTACTAGACCT 122

Db 283 TATCCCCCATGGAGACCTTCAGAAAGGTGGGATCCCCCATCATCATAGCACTACTAGACCT 342

QY 123 GGCAGTATCATCATTTGTTCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 182

Db 343 GGCAGTATCATCATTTGTTCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 402

QY 183 CTGGGCGAGCCTCTCCACTTCATCCGAGGAAGAGCTGTGTGACGAGAGCTGCACTG 242

Db 403 CTGGGCGAGCCTCTCCACTTCATCCGAGGAAGAGCTGTGTGACGAGAGCTGCACTG 462

QY 243 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCAGTGGC 302

Db 463 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCAGTGGC 522

QY 303 AGTCGGCTCTCCAGGACCGATCCACATCTGACAGTGTGGACTCGGCTCGGCACAGGAACTG 362

Db 523 AGTCGGCTCTCCAGGACCGATCCACATCTGACAGTGTGGACTCGGCTCGGCACAGGAACTG 582

QY 363 GTTCTCTGCTGTTCGACAACTTCACAGAACTCTCGCTGACAGCCTGTAGGCAGAT 422

Db 583 GTTCTCTGCTGTTCGACAACTTCACAGAACTCTCGCTGACAGCCTGTAGGCAGAT 642

QY 423 GGGCTACAGCAGCAACCCACTTTTCAGAGCTGTGGAGATTGGCCCAAGACAGGATCTGA 482

Db 643 GGGCTACAGCAGCAACCCACTTTTCAGAGCTGTGGAGATTGGCCCAAGACAGGATCTGA 702

QY 483 TGTGTGTAATCACAGAAACAGCCAGGAGCTTCGATCGCGAACTCAAGTGGGCTCTG 542

Db 703 TGTGTGTAATCACAGAAACAGCCAGGAGCTTCGATCGCGAACTCAAGTGGGCTCTG 762

QY 543 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTGTGCTGTGGGAAAGAGCTGAAGACCC 602

Db 763 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTGTGCTGTGGGAAAGAGCTGAAGACCC 822

QY 603 CCGTGTGGTGGTGGGAGGAGGAGCTCTGTGGATTTCTGGCCCTTGGCAGTCAAGTCCA 662

Db 823 CCGTGTGGTGGTGGGAGGAGGAGCTCTGTGGATTTCTGGCCCTTGGCAGTCAAGTCCA 882

QY 663 GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGACCCCACTGGGTCTCTACGGC 722

Db 883 GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTTGACCCCACTGGGTCTCTACGGC 942

QY 723 AGCCCACTGCTTCAGGAAACATACCCGATGTTCACCTGGAAGGTGGGCGAGGCTCAGA 782

Db 943 AGCCCACTGCTTCAGGAAACATACCCGATGTTCACCTGGAAGGTGGGCGAGGCTCAGA 1002

QY 783 CAAACTGGGAGCTTCCCATCTCCCTGGCTGTGGGCAAGATCATCATATTGAATCAACCC 842

Db 1003 CAAACTGGGAGCTTCCCATCTCCCTGGCTGTGGGCAAGATCATCATATTGAATCAACCC 1062

QY 843 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTCGAGTCCCACTCACTTTCTC 902

Db 1063 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTCGAGTCCCACTCACTTTCTC 1122

QY 903 AGGCACAGTCAAGGCCATCTGTCTGCCCTCTTTTGTGAGGAGCTCACTCCAGCCACCCC 962

Db 1123 AGGCACAGTCAAGGCCATCTGTCTGCCCTCTTTTGTGAGGAGCTCACTCCAGCCACCCC 1182

QY 963 ACTCTGATCATTTGGATGGGCTTTACGAAGCAATGAGAGGGAAGATGTCTGACATCT 1022

Db 1183 ACTCTGATCATTTGGATGGGCTTTACGAAGCAATGAGAGGGAAGATGTCTGACATCT 1242

QY 1023 GCTCGAGCGTCACTCCAGGTCATTGACGACACAGCTGCAATGACAGCAGTGGTACCA 1082

Db 1243 GCTCGAGCGTCACTCCAGGTCATTGAGCACAACGTTGCAATGACAGCAGTGGTACCA 1302

QY 1083 GGGGGAAGTCAACGAGAGATGTGTGAGGATCCCGGAAGGGGGTGTGACACCTG 1442

Db 1303 GGGGGAAGTCAACGAGAGATGTGTGAGGATCCCGGAAGGGGGTGTGACACCTG 1362

QY 1143 CCAGGGTGACAGTGGTGGCCCTCATGATACCAATCTGACAGTGGCATGTTGGTGGCAT 1202
DB 1363 CCAGGGTGACAGTGGTGGCCCTCATGATACCAATCTGACAGTGGCATGTTGGTGGCAT 1422
QY 1203 COTTAGCTGGGCTATGGCTGGGGGGCCCGAGACACCCAGAGATATACCAAGGTCTC 1262
DB 1423 COTTAGCTGGGCTATGGCTGGGGGGCCCGAGACACCCAGAGATATACCAAGGTCTC 1482
QY 1263 AGCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG 1305
DB 1483 AGCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG 1525
RESULT 6
ABS76529
ID ABS76529 standard; cDNA; 2307 BP.
AC ABS76529;
XX
DT 11-DEC-2002 (first entry)
XX
DE cDNA encoding human ovarian cancer marker OV86.
XX
KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;
central nervous system disorder; bacterial meningitis; viral meningitis;
Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
brain herniation; inflammation; encephalitis; testicular disorder;
nontuberculous granulomatous orchitis; connective tissue disorder;
heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
histological type; carcinogenic; ovarian cancer marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200271928-A2.
XX
PD 19-SEP-2002.
XX
PF 14-MAR-2002; 2002WO-US007825.
XX
PR 14-MAR-2001; 2001US-0276025P.
PR 14-MAR-2001; 2001US-0276026P.
PR 10-AUG-2001; 2001US-0311732P.
PR 19-SEP-2001; 2001US-0323580P.
PR 26-SEP-2001; 2001US-0324967P.
PR 26-SEP-2001; 2001US-0325102P.
PR 26-SEP-2001; 2001US-0325149P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
PI Meyers RE, Morrissey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
XX
DR WPI; 2002-723277/78.
DR P-PSDB; ABG96430.
XX
PT Assessing whether a patient is afflicted with ovarian cancer, useful in
PT assessing the stage or progression of the disease, comprises comparing
PT the expression level of a cancer marker in a sample from a patient and
PT from a non cancer patient.
XX
PS Disclosure; Page 438; 481pp; English.
XX
CC The present invention relates to a new method for assessing whether a
CC patient is afflicted with ovarian cancer. The method involves comparing
CC the expression level of a marker in a patient sample and the normal level
CC of expression of the marker in a control non-ovarian cancer sample, where
CC the marker is selected from 363 cancer markers described in the
CC specification. The method of the invention is useful in diagnosing or
CC characterising cancer, in detecting the presence of cancer as early as
CC possible, and the recurrence of ovarian cancer. The method may also be of
CC particular use with patients having an enhanced risk of developing

ovarian cancer (e.g. patients having a familial history of ovarian
cancer). The cancer markers may be used in the management and treatment
of e.g. brain and central nervous system disorders (e.g. bacterial and
viral meningitis, Alzheimer's disease or Parkinson's disease), brain
disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
inflammations (e.g. bacterial or viral meningitis or encephalitis),
testicular disorders (e.g. nontuberculous granulomatous orchitis),
connective tissue disorders, or heart disorders (e.g. ischaemic heart
disease or atherosclerosis). The compositions and methods may also be
used in assessing the histological type of neoplasm associated with
ovarian cancer, monitoring the progression of ovarian cancer, determining
whether ovarian cancer has metastasized or is likely to metastasize,
selecting a composition for inhibiting ovarian cancer, assessing the
ovarian carcinogenic potential of a compound, or inhibiting ovarian
cancer or at risk of developing ovarian cancer. The present nucleic acid
sequence encodes one of the ovarian cancer markers described in the
invention.
XX
SQ Sequence 2307 BP; 555 A; 647 C; 614 G; 460 T; 0 U; 31 Other;
Query Match 99.8%; Score 1303; DB 6; Length 2307;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATCTCTGACAGTCACTCACTCTGACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 62
DB 283 GGATCTCTGACAGTCACTCTGACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 342
QY 63 TATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCCATCATCATGACTACTAGCCCT 122
DB 343 TATCCCCATGGAGACCTTCAGAAAGTGGGGATCCCCATCATCATGACTACTAGCCCT 402
QY 123 GGGAGTATCATCATGTGTGTCTCTCATCAAGTGAATTCGGATTAATACTACTTCCT 182
DB 403 GGGAGTATCATCATGTGTGTCTCTCATCAAGTGAATTCGGATTAATACTACTTCCT 462
QY 183 CTGGGGCAGCCCTCTCCACTTTCACAGGAGAGCAGCTGTGACGAGAGTGGACTG 242
DB 463 CTGGGGCAGCCCTCTCCACTTTCACAGGAGAGCAGCTGTGACGAGAGTGGACTG 522
QY 243 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGGCTGCAATGGC 302
DB 523 TCCCTTGGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGGCTGCAATGGC 582
QY 303 AGTCCGCTCTCCAAAGGACCGATCCACACTGCGAGTGTGCGCCACAGGGAATCTG 362
DB 583 AGTCCGCTCTCCAAAGGACCGATCCACACTGCGAGTGTGCGCCACAGGGAATCTG 642
QY 363 GTTCTCTGCTGTGTTTCGACAACTTCAGAGAGCTTCTGCTGAGACAGCTGTAGGCAGAT 422
DB 643 GTTCTCTGCTGTGTTTCGACAACTTCAGAGAGCTTCTGCTGAGACAGCTGTAGGCAGAT 702
QY 423 GGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 482
DB 703 GGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 762
QY 483 TGTGTTGTAATCACAGAAACAGCCAGGAGCTTCGATCGGAATCTCAAGTGGGCCCTG 542
DB 763 TGTGTTGTAATCACAGAAACAGCCAGGAGCTTCGATCGGAATCTCAAGTGGGCCCTG 822
QY 543 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTTGTGCTGTGGGAAGAGCTGAAGACCC 602
DB 823 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTTGTGCTGTGGGAAGAGCTGAAGACCC 882
QY 603 CCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTTCTTGGCTTGGCAGGTGAGCATCCA 662
DB 883 CCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTTCTTGGCTTGGCAGGTGAGCATCCA 942
QY 663 GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCACTGGGTCTCTACGGC 722
DB 943 GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTTGGACCCCACTGGGTCTCTACGGC 1002
QY 723 AGCCCACTGCTTCAGGAACATACCCGATGTGTCAACTGGAGGTGGGGCAGGCTCAGA 782

Db 1003 AGCCCCACTCTTCAGGAACATACCGATGTTCACTGGAAGGTGCGGCGAGGCTCAGA 1062
 QY 783 CAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAGGATCATCATCAATGTAATCAACCC 842
 Db 1063 CAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCAGGATCATCATCAATGTAATCAACCC 1122
 QY 843 CATGTATCCCAAGACATGATGCGCTCATGAAGCTGAGTCCCACTCACTTTCTC 902
 Db 1123 CATGTATCCCAAGACATGATGCGCTCATGAAGCTGAGTCCCACTCACTTTCTC 1182
 QY 903 AGGCACAGTCAGGCCCATCTGCTGCTCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 962
 Db 1183 AGGCACAGTCAGGCCCATCTGCTGCTCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 1242
 QY 963 ACTGTGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATCT 1022
 Db 1243 ACTGTGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATCT 1302
 QY 1023 GCTGACGGCTGAGTCCAGTCAATGACACACACAGTGCATGACAGGATGCGTACCA 1082
 Db 1303 GCTGACGGCTGAGTCCAGTCAATGACACACACAGTGCATGACAGGATGCGTACCA 1362
 QY 1083 GGGGGAAGTCACCGAAGATGATGTGTGACGGCATCCCGGAAGGGGTGTGACACCTG 1142
 Db 1363 GGGGGAAGTCACCGAAGATGATGTGTGACGGCATCCCGGAAGGGGTGTGACACCTG 1422
 QY 1143 CAGGCTGACAGTGTGGGCTTGTGACGGCATCCCGGAAGGGGTGTGACACCTG 1202
 Db 1423 CAGGCTGACAGTGTGGGCTTGTGACGGCATCCCGGAAGGGGTGTGACACCTG 1482
 QY 1203 CGTTAGCTGGGCTATGCTGCTGGGCTTGTGACGGCATCCCGGAAGGGGTGTGACACCTG 1262
 Db 1483 CGTTAGCTGGGCTATGCTGCTGGGCTTGTGACGGCATCCCGGAAGGGGTGTGACACCTG 1542
 QY 1263 AGCTATCTCACTGATCTACATGCTGGAAGGCTGAGCTG 1305
 Db 1543 AGCTATCTCACTGATCTACATGCTGGAAGGCTGAGCTG 1585

RESULT 7

ACF12937
 ID ACF12937 standard; cDNA; 2307 BP.
 AC12937;
 10-SEP-2003 (first entry)
 Human cervical cancer cell marker encoding cDNA SEQ ID NO:217.
 Human; cervical cancer; cervical cancer marker; cancer therapy;
 detection; gene therapy; vaccine; gene; ss.
 Homo sapiens.
 WO2002101075-A2.
 19-DEC-2002.
 12-JUN-2002; 2002WO-US018638.
 13-JUN-2001; 2001US-0298155P.
 13-JUN-2001; 2001US-0298159P.
 14-NOV-2001; 2001US-0335936P.
 (MILL-) MILLENNIUM PHARM INC.
 Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;
 Gannavarapu M, Glatk K, Heersch S;
 WPI; 2003-156967/15.
 P-PSDB; ABR92154.

PT New isolated nucleic acid molecule useful for detecting, characterizing,
 preventing and treating human cervical cancers, in various prognostic and
 diagnostic assays, in pharmacogenomics and in monitoring clinical trials.
 XX
 PS Claim 1; Page 369; 386pp; English.
 XX

CC ACF12828 to ACF12947 encode the human cervical cancer marker proteins (I)
 given in ABR92164. A higher level of expression of (I) than
 normal indicates the presence of cervical cancer. Also described: (1) a
 vector (II) containing (I); (2) a host cell (III) containing (I); and (3)
 assessing (M1) whether a patient is afflicted with cervical cancer,
 comprising comparing the level of expression of a marker in a patient's
 sample, and the normal level of expression of the marker in a control non
 -cervical cancer sample, where a significant increase in the level of
 expression of the marker in the patient's sample relative to that in the
 control sample is an indication that the patient is afflicted with
 cervical cancer. (I) has cytostatic activity, and can be used in gene
 therapy and in vaccines. (I) is useful in detecting, characterizing,
 preventing and treating human cervical cancers. (I) may also be used in
 various prognostic and diagnostic assays, pharmacogenomics and in
 monitoring clinical trials

SQ Sequence 2307 BP; 555 A; 647 C; 614 G; 460 T; 0 U; 31 Other;

Query Match 99.8%; Score 1303; DB 7; Length 2307;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GGATCTGACAGTATCAACCTCTGACAGCTTCAGATGTCAACCCCTCGGGAACCCCG 62
 Db 283 GGATCTGACAGTATCAACCTCTGACAGCTTCAGATGTCAACCCCTCGGGAACCCCG 342
 QY 63 TATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCCTACTGAGCCT 122
 Db 343 TATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCCTACTGAGCCT 402
 QY 123 GGGAGTATCATATGTTGTTCTCTCATCAAGGTGATCTGGATAAATACTACTTCT 182
 Db 403 GGGAGTATCATATGTTGTTCTCTCATCAAGGTGATCTGGATAAATACTACTTCT 462
 QY 183 CTGGGGCAGCTCTCCATTCATCCGAGAAAGAGCTGTGTGACGGAGAGCTGAGCTG 242
 Db 463 CTGGGGCAGCTCTCCATTCATCCGAGAAAGAGCTGTGTGACGGAGAGCTGAGCTG 522
 QY 243 TCCCTTGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTGGC 302
 Db 523 TCCCTTGGGGAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTGGC 582
 QY 303 AGTCCGCTCTCCAAGGACCGATCCACATGACAGTGTGGACTCGGCCACAGGAACTG 362
 Db 583 AGTCCGCTCTCCAAGGACCGATCCACATGACAGTGTGGACTCGGCCACAGGAACTG 642
 QY 363 GTTCTCTGCTGTTTCGACAACTTCAAGAGCTCTCGCTGACAGAGCTGTAGGAGAT 422
 Db 643 GTTCTCTGCTGTTTCGACAACTTCAAGAGCTCTCGCTGACAGAGCTGTAGGAGAT 702
 QY 423 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATGGCCACAGACCATCTGA 482
 Db 703 GGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATGGCCACAGACCATCTGA 762
 QY 483 TGTGTTGTAATCACAGAAACAGCAGGAGCTTCGCATGCGGAACTCAAGTGGGCTG 542
 Db 763 TGTGTTGTAATCACAGAAACAGCAGGAGCTTCGCATGCGGAACTCAAGTGGGCTG 822
 QY 543 TCTCTAGGCTCCCTGTTCTCCCTGACATGTTCTGCTGTGGAAAGAGCTGAGACCCC 602
 Db 823 TCTCTAGGCTCCCTGTTCTCCCTGACATGTTCTGCTGTGGAAAGAGCTGAGACCCC 882
 QY 603 CCGTGTGGTGGGAGGAGGAGCTCTGTGGATTTCTGGCTTGGCAGTTCAGCATCCA 662
 Db 883 CCGTGTGGTGGGAGGAGGAGCTCTGTGGATTTCTGGCTTGGCAGTTCAGCATCCA 942
 QY 663 GTACGCAAAACAGCAGCTGTGTGGAGGAGCATCTCTGGACCCCACTGGTCTCTCAGGC 722

Db 943 GTACACAAACAGCAGCTCTGTGGAGGAGATCTTGAGCCCACTGGTCTCTACGGC 1002
Qy 723 AGCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 782
Db 1003 AGCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGTGGCGGAGGCTCAGA 1062
Qy 783 CAACTGGGAGCTCCCATCCCTGGCTGTGGCCNAGATCATCATGAATTAACCC 842
Db 1063 CAACTGGGAGCTCCCATCCCTGGCTGTGGCCNAGATCATCATGAATTAACCC 1122
Qy 843 CATGTACCCCAAGACATACCATCCCTGTATGAAGCTGCAGTTCCTCCCACTCTTCTC 902
Db 1123 CATGTACCCCAAGACATACCATCCCTGTATGAAGCTGCAGTTCCTCCCACTCTTCTC 1182
Qy 903 AGGCACAGTCCAGGCTCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 962
Db 1183 AGGCACAGTCCAGGCTCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1242
Qy 963 ACTCTGGATCATGTGATGGGCTTTACGAAGCAGAAATGAGGAGATGTCTGCATACT 1022
Db 1243 ACTCTGGATCATGTGATGGGCTTTACGAAGCAGAAATGAGGAGATGTCTGCATACT 1302
Qy 1023 GCTGAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGACAGATGGGTACCA 1082
Db 1303 GCTGAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCATGACAGATGGGTACCA 1362
Qy 1083 GGGGAAGTCCAGGAGATGATGTGTGAGGATCCCGGAGGCTGTGGACCTG 1142
Db 1363 GGGGAAGTCCAGGAGATGATGTGTGAGGATCCCGGAGGCTGTGGACCTG 1422
Qy 1143 CCAGGTCACAGTGTGGGCTCTGATGACCAATCTGACAGTGGCATGTGTGGGCT 1202
Db 1423 CCAGGTCACAGTGTGGGCTCTGATGACCAATCTGACAGTGGCATGTGTGGGCT 1482
Qy 1203 CTTAGTGGGCTATGCTGGGCGCCGAGACCCGAGAGTATACCAAGTCTC 1262
Db 1483 CTTAGTGGGCTATGCTGGGCGCCGAGACCCGAGAGTATACCAAGTCTC 1542
Qy 1263 AGCCTATCTCAACTGATCTCAATGTCTGGAAGCTGAGCTG 1305
Db 1543 AGCCTATCTCAACTGATCTCAATGTCTGGAAGCTGAGCTG 1585
RESULT 8
AAI64284
ID AAI64284 standard; cDNA; 2121 BP.
XX AAI64284;
AC AAI64284;
XX 07-MAY-2002 (first entry)
DT Human serine protease D-G cDNA.
DE Serine protease; D-G; human; zymogen; enzyme; cytostatic;
XX antinflammatory; dermatological; anticoagulation; cancer; skin disorder;
KW neuropathic pain; inflammatory disorder; coagulation diathesis;
KW thrombosis; laundry detergent; skin care; gene therapy; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 277..1584
CDS /*tag= a
FT /*product= "serine protease D-G"
FT /*transl_except= (pos: 1357..1359, aa: Gln)
FT polyA_signal 2100..2105
FT /*tag= b
FT /*notes= "does not conform to consensus sequence"
XX W0200202011-A1.
PN 10-JAN-2002.
XX
PD

XX 08-JUN-2001; 2001WO-US018568.
XX 30-JUN-2000; 2000US-00607745.
XX (ORTH) ORTHO-MCNEIL PHARM INC.
XX Darrow AL, Qi J, Andrade-Gordon P;
XX WPI; 2002-106601/14.
XX P-PSDB; AAG78577.
XX Nucleic acid encoding a serine protease called D-G protein which is
XX useful for identifying modulators that are useful for treating a
XX condition which is mediated by protease D-G, e.g. cancer, skin disorders,
XX or neuropathic pain.
XX Claim 4; Fig 1A; 81pp; English.
XX The invention relates to an isolated and purified nucleic acid that
XX encodes a serine protease called D-G protein. The activity of the protein
XX of the invention may be described as cytostatic, antinflammatory,
XX dermatological and anticoagulation. The serine protease of the invention
XX is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
XX which play an important role in processes such as digestion and
XX regulatory amplification cascades through the proteolytic activation of
XX inactive zymogen precursors. Protease D-G modulating compounds are useful
XX for treating a condition which is mediated by protease D-G, e.g. cancer,
XX skin disorders, neuropathic pain, inflammatory disorders, or coagulation
XX diathesis/thrombosis. The polynucleotide encoding the protease is useful
XX for identifying, detecting or isolating mutant forms of DNA molecules
XX encoding the protease. The protease is useful for identifying modulators
XX of the functional protease. The D-G protein can be used for formulation
XX of compositions for laundry detergents and skin care products. Protease D
XX -G gene therapy may be used to introduce protease D-G into the cells of
XX target organisms. As the D-G protein is derived from a human, it is less
XX likely to produce an allergic reaction in sensitive individuals when used
XX in formulations for laundry detergents and skin care products. The
XX current sequence represents the human serine protease D-G cDNA
XX Sequence 2121 BP; 487 A; 614 C; 584 G; 436 T; 0 U; 0 Other;
Query Match 99.8%; Score 1301.8; DB 6; Length 2121;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGATCCTGACAGTGTCAACCTCTGAACAGGCTCGATGTCAAAACCCCTGGCAACCC 60
Db 277 ATGATCCTGACAGTGTCAACCTCTGAACAGGCTCGATGTCAAAACCCCTGGCAACCC 336
Qy 61 CGTATCCCAATGGAGACCTTCAGAAAGTGGGATCCCATCATCATGACTACTGAGC 120
Db 337 CGTATCCCAATGGAGACCTTCAGAAAGTGGGATCCCATCATCATGACTACTGAGC 396
Qy 121 CTGCGAGTATCATCTGTGTGTCTCTCATCAAGGTGATCTCTGGAATAAATACTTCTC 180
Db 397 CTGCGAGTATCATCTGTGTGTCTCTCATCAAGGTGATCTCTGGAATAAATACTTCTC 456
Qy 181 CTCTCGGGGAGGCTCTCCACTTTCATCCGAGGAGGAGCTGTGTGACGAGAGCTGGAC 240
Db 457 CTCTCGGGGAGGCTCTCCACTTTCATCCGAGGAGGAGCTGTGTGACGAGAGCTGGAC 516
Qy 241 TGTCCTTGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAGGCTTGGAGT 300
Db 517 TGTCCCTTGGGGAGGAGGAGGAGCTGTGTCAAGAGCTTCCCGAGGCTTGGAGT 576
Qy 301 GCAGTCCGCTCTCTCAAGGACCGATCCACACTGCAAGTGTGTGACTCGGCCACAGGAAC 360
Db 577 GCAGTCCGCTCTCTCAAGGACCGATCCACACTGCAAGTGTGTGACTCGGCCACAGGAAC 636
Qy 361 TGGTCTCTGCTGTCTTTCGACAACTTCACAGAGCTCTCGCTGAGAGCTGTAGCAG 420
Db 637 TGGTCTCTGCTGTCTTTCGACAACTTCACAGAGCTCTCGCTGAGAGCTGTAGCAG 696

QY 421 ATGGGCTACAGCAGCAAAACCCATTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 480
 DB 697 ATGGGCTACAGCAGCAAAACCCATTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 756
 QY 481 GATGTTGTTGAATACAGAAACACCCAGGAGCTTCGATCGGNACTCAAGTGGGCC 540
 DB 757 GATGTTGTTGAATACAGAAACACCCAGGAGCTTCGATCGGNACTCAAGTGGGCC 816
 QY 541 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACCTGTCTTCCCTGTGGGAAGAGCCTGAAGACC 600
 DB 817 TGTCTCTCAGGCTCCCTGCTCTCCCTGCACCTGTCTTCCCTGTGGGAAGAGCCTGAAGACC 876
 QY 601 CCCGTTGTTGGTGGGAGGAGGCTCTGTGGATCTTTGGCTTTGGAGGTTCAGATC 660
 DB 877 CCCGTTGTTGGTGGGAGGAGGCTCTGTGGATCTTTGGCTTTGGAGGTTCAGATC 936
 QY 661 CAGTACGACAAACAGCAGCAGCTGTGTGGAGGAGCATCTCGACCCGCCACCTGCTCAGG 720
 DB 937 CAGTACGACAAACAGCAGCAGCTGTGTGGAGGAGCATCTCGACCCGCCACCTGCTCAGG 996
 QY 721 GCAGCCCACTGCTTCAGGAAACATACCGATGTTCACACTGGAAGGTGCGGCAGGCTCA 780
 DB 997 GCAGCCCACTGCTTCAGGAAACATACCGATGTTCACACTGGAAGGTGCGGCAGGCTCA 1056
 QY 781 GACAACTGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCAATCAAC 840
 DB 1057 GACAACTGGCAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCAATCAAC 1116
 QY 841 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTTC 900
 DB 1117 CCATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGCAGTTCCTCACTTTC 1176
 QY 901 TCAGGCAGAGTCAGGCCCACTGTCTGCCCTTCTTTGATGAGAGCTCACTCAGCCACC 960
 DB 1177 TCAGGCAGAGTCAGGCCCACTGTCTGCCCTTCTTTGATGAGAGCTCACTCAGCCACC 1236
 QY 961 CCACTCTGATCATTCGATGGGGCTTTACAGAGCAAGTGGAGGAAGATGCTGACATA 1020
 DB 1237 CCACTCTGATCATTCGATGGGGCTTTACAGAGCAAGTGGAGGAAGATGCTGACATA 1296
 QY 1021 CTGCTCAGCGGTGAGTCAGGTCAATGACAGACACCGTGCATGCGAGACGATCGGTAC 1080
 DB 1297 CTGCTCAGCGGTGAGTCAGGTCAATGACAGACACCGTGCATGCGAGACGATCGGTAC 1356
 QY 1081 CAGGGGGAAGTCAACGAGAGATGATGTGAGGACATCCCGAAGGGGTGTGGACACC 1140
 DB 1357 CTGGGGGAAGTCAACGAGAGATGATGTGAGGACATCCCGAAGGGGTGTGGACACC 1416
 QY 1141 TGCCAGGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACCAAGTGCATGTGGTGGGC 1200
 DB 1417 TGCCAGGGTGACAGTGGTGGGCCCTTGATGTACCAATCTGACCAAGTGCATGTGGTGGGC 1476
 QY 1201 ATCGTTAGCTGGGGCTATGGCTGCGGGGCCCGAGCAGCCAGGAGTATACACAGGTC 1260
 DB 1477 ATCGTTAGCTGGGGCTATGGCTGCGGGGCCCGAGCAGCCAGGAGTATACACAGGTC 1536
 QY 1261 TCAGCCATCTCAACTGGAATCTACAAATGTCTGGAAGGCTGAGCTG 1305
 DB 1537 TCAGCCATCTCAACTGGAATCTACAAATGTCTGGAAGGCTGAGCTG 1581

RESULT 9

AADI3114

ID AADI3114 standard; DNA; 2137 BP.

XX

AC AADI3114;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human membrane-type serine protease (MTSP) 3 DNA.

XX

KW Human; transmembrane serine protease; membrane-type serine protease;

MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
 lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
 ds.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 261..1574
 FT /*tag= a
 FT /product= "Human transmembrane serine protease (MTSP) 3"
 PN W0200157194-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 02-FEB-2001; 2001WO-US003471.
 XX
 XX 03-FEB-2000; 2000US-0179982P.
 PR 18-FEB-2000; 2000US-0183542P.
 PR 22-JUN-2000; 2000US-0213124P.
 PR 26-JUL-2000; 2000US-0220970P.
 PR 08-SEP-2000; 2000US-00657986.
 PR 22-SEP-2000; 2000US-0234840P.
 XX
 PA (CORV-) CORVAS INT INC.

Madison EL, Ong EO, Yeh J;
 WPI; 2001-488877/53.
 P-PSDB; AAE06931.
 Novel single chain polypeptide comprising protease domain of type-II
 membrane-type serine protease or its catalytically active portion useful
 for treating and preventing cancer and tumor.
 Claim 14; Page 197-199; 256pp; English.

The invention relates to transmembrane serine proteases and their
 corresponding nucleotides and the protease domain of a type-II membrane-
 type serine protease (MTSP). MTSP is useful for identifying compounds
 that modulate or inhibit its proteolytic activity and for formulating a
 medicament for treating neoplastic disease. MTSP and its corresponding
 nucleotides are useful in preventing or treating tumours or cancers such
 as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in
 diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
 marker for tumour development, growth and/or progression and as
 immunogens to generate antibodies that specifically bind to it. MTSP DNA
 is useful in a yeast two-hybrid system and in gene therapy. The present
 sequence is a DNA encoding human MTSP3 protein.

Sequence 2137 BP; 508 A; 612 C; 589 G; 428 T; 0 U; 0 Other;

Query Match

Best Local Similarity 99.5%; Score 1298.2; DB 4; Length 2137;

Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GGATCTCGACAGTCATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTCGCAAAACCCCG 62
 DB 269 GGATCTCGACAGTCATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTCGCAAAACCCCG 328
 QY 63 TATCCCCATGGAGACCTTCAGAAAGTGGGATCCCCCATCATAGCACTACTGAGCCT 122
 DB 329 TATCCCCATGGAGACCTTCAGAAAGTGGGATCCCCCATCATAGCACTACTGAGCCT 388
 QY 123 GCGAGTATCATATTGTGTCTCTCATCAAGGTGATTCTCGATAAATACTACTTCTCT 182
 DB 389 GCGAGTATCATATTGTGTCTCTCATCAAGGTGATTCTCGATAAATACTACTTCTCT 448
 QY 183 CTGCGGCGAGCCTCTCCACTTTCATCCGAGGAGAGAGCTGTGTGACGAGAGCTGGACTG 242
 DB 449 CTGCGGCGAGCCTCTCCACTTTCATCCGAGGAGAGAGCTGTGTGACGAGAGCTGGACTG 508
 QY 243 TCCTTTGGGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCCGAGGCGCTGCAGTGGC 302

SQ Sequence 2165 BP; 502 A; 623 C; 595 G; 445 T; 0 U; 0 Other;

Query Match 99.2%; Score 1294; DB 6; Length 2165;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGATCTGACAGTATCAACCTCTGAAACGCTCGATGTCAAACCCCTGCGCAAAACCC 60
DB 310 ATGGATCTGACAGTATCAACCTCTGAAACGCTCGATGTCAAACCCCTGCGCAAAACCC 369

QY 61 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 370 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 429

QY 121 CTGGGAGTATCATCATTTGGTGTCTCTCATCAAGTGATTCGGATAAATACTACTTC 180
DB 430 CTGGGAGTATCATCATTTGGTGTCTCTCATCAAGTGATTCGGATAAATACTACTTC 489

QY 181 CTCTCGGGGACGCTCTCCACTTCATCCGAGGACAGCTGTGTGAGCGGAGCTGGAC 240
DB 490 CTCTCGGGGACGCTCTCCACTTCATCCGAGGAGAGCTGTGTGAGCGGAGCTGGAC 549

QY 241 TGTCCCTTGGGGAGGACGAGGACGACTGTGTCAAGACTTCCCGAGGCGCTGCAGTG 300
DB 550 TGTCCCTTGGGGAGGACGAGGACGACTGTGTCAAGACTTCCCGAGGCGCTGCAGTG 609

QY 301 GCAGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGGAAC 360
DB 610 GCAGTCCGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGGAAC 669

QY 361 TGGTTCTCTGCTGTTTCGAACTTCAGAACTTCAGAACTTCGCTGAGACGACCTGTAGGAG 420
DB 670 TGGTTCTCTGCTGTTTCGAACTTCAGAACTTCAGAACTTCGCTGAGACGACCTGTAGGAG 729

QY 421 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 480
DB 730 ATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGCAGGATCTG 789

QY 481 GATGTTGTGAAATCACAGAAACAGCCAGGA - GCTTCGCAATGCGGAACCTCAAGTGGGCC 539
DB 790 GATGTTGTGAAATCACAGAAACAGCCAGGAGCTTCGCAATGCGGAACCTCAAGTGGGCC 849

QY 540 CTGTTCTCTCAGGCTCCCTGCTCTCCCTGACCTGCTTCCCTGCTGGAGAGCCTGAGAC 599
DB 850 CTGTTCTCTCAGGCTCCCTGCTCTCCCTGACCTGCTTCCCTGCTGGAGAGCCTGAGAC 909

QY 600 CCCCCGCTGTGGTGGGGAGGAGGCTCTGTGGAATCTTGGCTTGGCAGGTGAGCAT 659
DB 910 CCCCCGCTGTGGTGGGGAGGAGGCTCTGTGGAATCTTGGCTTGGCAGGTGAGCAT 969

QY 660 CCAGTACGACAAACAGCAGCTGTGGAGGAGGACATCTGACACCCCACTGGTCTCTCAC 719
DB 970 CCAGTACGACAAACAGCAGCTGTGGAGGAGGACATCTGACACCCCACTGGTCTCTCAC 1029

QY 720 GGCAGCCCACTCTTCAGAAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTC 779
DB 1030 GGCAGCCCACTCTTCAGAAACATACCGATGTGTTCAACTGGAGGTGGGGCAGGCTC 1089

QY 780 AGACAAATGGGAGCTTCCCAATCCCTGCTGTGGCCAAAGATCATCATTTGATTTCAA 839
DB 1090 AGACAAATGGGAGCTTCCCAATCCCTGCTGTGGCCAAAGATCATCATTTGATTTCAA 1149

QY 840 CCCCATGTACCCCAAGCAATGACATCCCTCATGAGCTGCAAGTCCCACTCACTTT 899
DB 1150 CCCCATGTACCCCAAGCAATGACATCCCTCATGAGCTGCAAGTCCCACTCACTTT 1209

QY 900 CTCAGGACAGTCAAGGCCCATCTGTCTGCCCTTTCTTTGATGAGGAGCTCACTCCAGCCAC 959
DB 1210 CTCAGGACAGTCAAGGCCCATCTGTCTGCCCTTTCTTTGATGAGGAGCTCACTCCAGCCAC 1269

QY 960 CCCACTCTGGATCATTTGATGGGCTTTTACGAGCAGATGAGGAGGAGATGCTTGACAT 1019
DB 1270 CCCACTCTGGATCATTTGATGGGCTTTTACGAGCAGATGAGGAGGAGATGCTTGACAT 1329

QY 1020 ACTGCTGCAGGCGTCCAGTCCAGTCAATTCAGACGACACGCTGCAATGCAGATCGGTA 1079
DB 1330 ACTGCTGCAGGCGTCCAGTCCAGTCAATTCAGACGACACGCTGCAATGCAGATCGGTA 1389

QY 1080 CCAGGGGGAAGTCCACGAGAAGATGATGTGTCAGGATATCCCGGAAGGGGTGTGGACAC 1139
DB 1390 CCAGGGGGAAGTCCACGAGAAGATGATGTGTCAGGATATCCCGGAAGGGGTGTGGACAC 1449

QY 1140 CTGCGAGGTGACAGTGTGGGCGCTGATGTACCAATCTGACCACTGCAATGTGGTGG 1199
DB 1450 CTGCGAGGTGACAGTGTGGGCGCTGATGTACCAATCTGACCACTGCAATGTGGTGG 1509

QY 1200 CATCTGTTAGTGGGCTATGCTGTCGGGGCGCCGAGCACCCAGGAGTATACACCAAGGT 1259
DB 1510 CATCTGTTAGTGGGCTATGCTGTCGGGGCGCCGAGCACCCAGGAGTATACACCAAGGT 1569

QY 1260 CTGAGCTATCTCAACTGATCTCAATGTCTGGAAGGCTGAGCTG 1305
DB 1570 CTGAGCTATCTCAACTGATCTCAATGTCTGGAAGGCTGAGCTG 1615

RESULT 12
AAZ90471
ID AAZ90471 standard; cDNA; 2070 BP.
XX AAZ90471;
XX AC
XX AC
XX 06-JUN-2000 (first entry)
XX
DE Cancer specific gene (CSG) sequence (clone ID 1283171).
XX CSG; cancer specific gene; cancer; gynecologic cancer; ovarian; breast;
XX endometrial; uterine; lung; cytotoxic; ss.
XX Homo sapiens.
XX
XX WO200012758-A1.
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US019655.
XX
XX 02-SEP-1998; 98US-0098880P.
XX
XX (DIAD-) DIADEXUS LLC.
XX
XX Salceda S, Sun Y, Recipon H, Cafferkey R;
XX WPI; 2000-256657/22.
XX
XX Diagnosing, staging, monitoring, imaging and treating cancer especially
XX gynecological cancers e.g. breast, ovarian cancer and lung cancer,
XX involves measuring cancer specific gene levels in cells and body fluids.
XX
XX Claim 9; Page 45-46; 58pp; English.
XX
XX The invention relates to detecting, diagnosing metastasis and staging
XX cancer by measuring levels of cancer specific genes (CSG) in cells,
XX tissues or body fluids. Their remission and progression, decreases and
XX increases in CSG levels, is also monitored, by periodic sample analysis.
XX The methods are useful for detecting cancers, especially gynecologic
XX cancers which include ovarian, breast, endometrial and uterine cancer and
XX lung cancer. Antibodies against the CSGs labeled with paramagnetic ions
XX or a radioisotope is useful for imaging cancer and when conjugated with a
XX cytotoxic agent are useful for treating cancer. The present sequence
XX represents a CSG sequence (clone ID: 1283171 and gene ID: 332459)
XX
SQ Sequence 2070 BP; 473 A; 599 C; 577 G; 421 T; 0 U; 0 Other;

Query Match 99.0%; Score 1292; DB 3; Length 2070;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 GGATCTGTGAGTCAACCTCTGAACAGCTCGATGTCAAACCCCTGGCGAAACCCCG 62
DB 222 GGATCTGTGAGTCAACCTCTGAACAGCTCGATGTCAAACCCCTGGCGAAACCCCG 281
QY 63 TATCCCATGAGAGCTTCAGAAAGTGGGATGCCCATCATAGCACTACTGAGCCT 122
DB 282 TATCCCATGAGAGCTTCAGAAAGTGGGATGCCCATCATAGCACTACTGAGCCT 341
QY 123 GCGAGTATCATCTGTTGTTCTCATCAAGGTGATCTCGAATAAATACTACTTCT 182
DB 342 GCGAGTATCATCTGTTGTTCTCATCAAGGTGATCTCGAATAAATACTACTTCT 401
QY 183 CTGGCGGAGCCTCTCCACTTCATCCGAGGAAGAGCTGTGTGACGAGAGCTGGACTG 242
DB 402 CTGGCGGAGCCTCTCCACTTCATCCGAGGAAGAGCTGTGTGACGAGAGCTGGACTG 461
QY 243 TCCCTTGGGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCACTGGC 302
DB 462 TCCCTTGGGGAGGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCGCTGCACTGGC 521
QY 303 AGTCGCGCTCTCCAGAGCCGATCCACTGCACTGCACTGCACTGCACTGCACTGCACTG 362
DB 522 AGTCGCGCTCTCCAGAGCCGATCCACTGCACTGCACTGCACTGCACTGCACTGCACTG 581
QY 363 GTTCTCTGCTGTTGCAAACTTCACAGAGCTCTCGCTGAGAGAGCTGTGAGCAGAT 422
DB 582 GTTCTCTGCTGTTGCAAACTTCACAGAGCTCTCGCTGAGAGAGCTGTGAGCAGAT 641
QY 423 GGGCTACAGCAGCAAAACCTTTTCAGAGCTGTGTGAGATTTGCCAGACAGGATCTGGA 482
DB 642 GGGCTACAGCAGCAAAACCTTTTCAGAGCTGTGTGAGATTTGCCAGACAGGATCTGGA 701
QY 483 TGTGTGAAATCACAGAAAAAGCAGGAGCTTCGCATGCGGAATCAAGTGGGCGCTG 542
DB 702 TGTGTGAAATCACAGAAAAAGCAGGAGCTTCGCATGCGGAATCAAGTGGGCGCTG 761
QY 543 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
DB 762 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 821
QY 603 CCGTGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTTGGCAGGTGAGCATCCA 662
DB 822 CCGTGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTTGGCAGGTGAGCATCCA 881
QY 663 GTACGACAAAGCAGCTGTGTGAGGAGAGCATCTGCAACCCCTGCTGCTGCTGCTGCTG 721
DB 882 GTACGACAAAGCAGCTGTGTGAGGAGAGCATCTGCAACCCCTGCTGCTGCTGCTGCTG 941
QY 722 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAG 781
DB 942 CAGCCCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAG 1001
QY 782 ACAAACTGGGAGCTTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
DB 1002 ACAAACTGGGAGCTTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1061
QY 842 CATGTATACCCAAAGACATGATGCGCTTCATGAAGCTGAGTCCCATCTCACTTTCT 901
DB 1062 CATGTATACCCAAAGACATGATGCGCTTCATGAAGCTGAGTCCCATCTCACTTTCT 1121
QY 902 CAGGCACAGTCAGGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
DB 1122 CAGGCACAGTCAGGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1181
QY 962 CACTCTGATCATTTGATGGGCTTTAGAGCAATGGAGGAAGATGCTGACATAC 1021
DB 1182 CACTCTGATCATTTGATGGGCTTTAGAGCAATGGAGGAAGATGCTGACATAC 1241
QY 1022 TGCTCAGGCGTCACTCCAGGTCAATGACAGCAGCAGTGAATGCAACGATCGTACC 1081
DB 1242 TGCTCAGGCGTCACTCCAGGTCAATGACAGCAGCAGTGAATGCAACGATCGTACC 1301

1082 AGGGGAAGTCAACGAGAGTATGTGTGAGGATCCCGAAGGGGTGTGGACACT 1141
1302 AGGGGAAGTCAACGAGAGTATGTGTGAGGATCCCGAAGGGGTGTGGACACT 1361
1142 GCAGAGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCA 1201
1362 GCAGAGTGACAGTGTGGGCCCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCA 1421
1202 TCGTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACCAAGGTCT 1261
1422 TCGTTAGCTGGGCTATGGCTGGGGGGCCCGAGCACCCCGAGGATATACCAAGGTCT 1481
1262 CAGCTATCTCAACTGGAATCAATGTCTGGAAGCTGAGCTG 1305
1482 CAGCTATCTCAACTGGAATCAATGTCTGGAAGCTGAGCTG 1525

RESULT 13
AAD13169
ID AAD13169 standard; DNA; 2079 BP.
XX
AC AAD13169;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human transmembrane protease serine 4 (TPRSS4) DNA.
XX
KW Human; transmembrane serine protease; membrane-type serine protease;
KW MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
KW lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
KW TPRS4; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 251..1522
FT /*tag= a
FT /product= "Human transmembrane protease, serine 4
FT (TPRSS4)"
XX
FN WO200157194-A2.
XX
PD 09-AUG-2001.
XX
PF 02-FEB-2001; 2001WO-US003471.
XX
PR 03-FEB-2000; 2000US-0179982P.
PR 18-FEB-2000; 2000US-0183542P.
PR 22-JUN-2000; 2000US-0213124P.
PR 26-JUL-2000; 2000US-0220970P.
PR 08-SEP-2000; 2000US-00657986.
PR 22-SEP-2000; 2000US-0234840P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Madison EL, Ong EO, Yeh J;
XX
XX WPI; 2001-488877/53.
XX P-PSDB; AAE06944.
XX
XX Novel single chain polypeptide comprising protease domain of type-II
XX membrane-type serine protease or its catalytically active portion useful
XX for treating and preventing cancer and tumor.
XX
XX Disclosure; Page 253-255; 256pp; English.
XX
XX The invention relates to transmembrane serine proteases and their
XX corresponding nucleotides and the protease domain of a type-II membrane-
XX type serine protease (MTSP). MTSP is useful for identifying compounds
XX that modulate or inhibit its proteolytic activity and for formulating a
XX medicament for treating neoplastic disease. MTSP and its corresponding
XX nucleotides are useful in preventing or treating tumours or cancers such
XX as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in

CC	diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
CC	marker for tumour development, growth and/or progression and as
CC	immunogens to generate antibodies that specifically bind to it. MTSP DNA
CC	is useful in a Yeast two-hybrid system and in gene therapy. The present
CC	sequence is a DNA encoding human transmembrane protease, serine 4
CC	(TMPRSS4) protein
XX	
SQ	Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 U; 0 Other;
	Query Match 99.0%; Score 1292; DB 4; Length 2079;
	Best Local Similarity 99.9%; Pred. No. 0;
	Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY	3 GGATCTGTGACAGTATCAACCTCTGTACACGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 62
DB	216 GGATCTGTGACAGTATCAACCTCTGTACACGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 275
QY	63 TATCCCATGGAGACCTTTCAGAAG-GTGGGATCCCATCATCATATAGCACTACTGAGCC 121
DB	276 TATCCCATGGAGACCTTTCAGAAGGTGGGATCCCATCATCATATAGCACTACTGAGCC 335
QY	122 TGGCGAGTATCATGTGGTGTCTCATCAAGGTATCTCGATAAATACTACTTCC 181
DB	336 TGGCGAGTATCATCAATGTGGTGTCTCATCAAGGTATCTCGATAAATACTACTTCC 395
QY	182 TCTGCGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGACTCGAAT 241
DB	396 TCTGCGGGCAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGACTCGAAT 455
QY	242 GTCCCTTGGGGAGGACGAGGACACTGTGTCAAGACTTCCCGAGGCGCTGCAGTGG 301
DB	456 GTCCCTTGGGGAGGACGAGGACACTGTGTCAAGACTTCCCGAGGCGCTGCAGTGG 515
QY	302 CAGTCCGCCTCTCCAAGGACCGATCCACACTGACAGTGTGGACTCGGCCACAGGGAAT 361
DB	516 CAGTCCGCCTCTCCAAGGACCGATCCACACTGACAGTGTGGACTCGGCCACAGGGAAT 575
QY	362 GGTTCCTCGCTGTTTGCACAACTTCACAGAACTCTCGCTGACAGCCTGTAGGCAGA 421
DB	576 GGTTCCTCGCTGTTTGCACAACTTCACAGAACTCTCGCTGACAGCCTGTAGGCAGA 635
QY	422 TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCCACAGCATCTGG 481
DB	636 TGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCCACAGCATCTGG 695
QY	482 ATGTTGTTGAAATCAGAGAAACAGCCAGAGCTTCGATTCGGAATCTCAAGTGGGCGCT 541
DB	696 ATGTTGTTGAAATCAGAGAAACAGCCAGAGCTTCGATTCGGAATCTCAAGTGGGCGCT 755
QY	542 GTCTCTCAGGCTCCCTGTGCTCCCTGCACCTGTCTGTGCTGTGGGAAAGAGCTGAAGACCC 601
DB	756 GTCTCTCAGGCTCCCTGTGCTCCCTGCACCTGTCTGTGCTGTGGGAAAGAGCTGAAGACCC 815
QY	602 CCCGTGTGGTGGGAGAGAGCCTCTGTGTGATTCCTTGGCCTTGGCAGGTGAGCATCC 661
DB	816 CCCGTGTGGTGGGAGAGAGCCTCTGTGTGATTCCTTGGCCTTGGCAGGTGAGCATCC 875
QY	662 AGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGTGACCCCCACTTGGTCTCTCAGG 721
DB	876 AGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGTGACCCCCACTTGGTCTCTCAGG 935
QY	722 CAGCCCATGTCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAG 781
DB	936 CAGCCCATGTCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAG 995
QY	782 ACAAACTGGGAGCTTCCCATCTCCCTGGCTGTGGCCACAGATCATCATTTGAAATCAACC 841
DB	996 ACAAACTGGGAGCTTCCCATCTCCCTGGCTGTGGCCACAGATCATCATTTGAAATCAACC 1055
QY	842 CCATGTATCCCAAAGACAAATGACATCGCCCTCATGAGCTGCAGTTCCTACTCATTTCT 901
DB	1056 CCATGTATCCCAAAGACAAATGACATCGCCCTCATGAGCTGCAGTTCCTACTCATTTCT 1115

Qy	902	CAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC	961
Db	1116	CAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCC	1175
Qy	962	CACCTCTGGATCATTTGGATGGGGCTTTACGAAGCAGCAATGGAGGAAGATGCTCTGACATAC	1021
Db	1176	CACCTCTGGATCATTTGGATGGGGCTTTACGAAGCAGCAATGGAGGAAGATGCTCTGACATAC	1235
Qy	1022	TGCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGCAGACGATCGGTACC	1081
Db	1236	TGCTGCAGGCGTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGCAGACGATCGGTACC	1295
Qy	1082	AGGGGGAAGTCACCCGAGAGATGATGTGTGAGGCATCCCGAAGGGGGTGTGGACACCT	1141
Db	1296	AGGGGGAAGTCACCCGAGAGATGATGTGTGAGGCATCCCGAAGGGGGTGTGGACACCT	1355
Qy	1142	GCCAGGCTGACAGTGTGGGCCCCCTGATGTACCAATCTGCACCATGTGGTGGGCA	1201
Db	1356	GCCAGGCTGACAGTGTGGGCCCCCTGATGTACCAATCTGCACCATGTGGTGGGCA	1415
Qy	1202	TGCTTACCTGGGCTATGCTGGGGGGCCCGAGCACCCGAGGATATACCAAGGTCT	1261
Db	1416	TCGTTAGCTGGGCTATGCTGGGGGGCCCGAGCACCCGAGGATATACCAAGGTCT	1475
Qy	1262	CAGCCTATCTCAACTGGATCTACAATGCTCTGGAAGGCTGAGCTG	1305
Db	1476	CAGCCTATCTCAACTGGATCTACAATGCTCTGGAAGGCTGAGCTG	1519
RESULT 14			
AAI72975			
ID	AAI72975 standard; cDNA; 2079 BP.		
XX			
AC	AAI72975;		
XX			
DT	21-AUG-2002 (first entry)		
XX			
DE	CJAB cDNA.		
XX			
KW	Gene; colorectal cancer; CGAV; CJAB; modulating protein; screening;		
KW	drug candidate; vaccine; ss.		
XX			
OS	Homo sapiens.		
XX			
Key	Location/Qualifiers		
FT	302..1522		
FT	/*tag= a		
FT	/product= "CJAB"		
XX			
FN	US2002042067-A1.		
XX			
PD	11-APR-2002.		
XX			
PF	08-MAY-2001; 2001US-00851588.		
XX			
PR	17-AUG-2000; 2000US-00642252.		
PR	06-SEP-2000; 2000US-00656002.		
XX			
PA	(MACK/) MACK D.		
PA	(GISH/) GISH K C.		
PA	(WILS/) WILSON K E.		
XX			
PI	Mack D, Gish KC, Wilson KE;		
XX			
DR	WPI; 2002-453647/48.		
DR	P-PSDB; AAG79358.		
XX			
PT	Screening drug candidates for treating colorectal cancer, comprises		
PT	determining the effect of the candidate on the expression profile gene of		
PT	CGAV or CJAB.		
XX			
PS	Disclosure; Fig 5; 40pb; English.		
XX			

XX screening drug candidates for treating colorectal cancer, comprises
PT determining the effect of the candidate on the expression profile gene of
PT CGA7 or CUA8.
XX
XX
PS Disclosure: Fig 5; 40pp; English.

The sequences given in AAI2973-76 encode the colorectal cancer proteins, CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal cancer modulating proteins and have been mapped to chromosomes 2 (CGA7) and 11 (CJA8). These sequences may be used in the method of the invention for screening drug candidates. The method comprises adding a drug candidate to a cell that expresses an expression profile gene encoding CGA7, CJA8 or a fragment and determining the effect of the drug candidate on the expression of the expression profile gene. The new methods are used to screen bioactive agents for the ability to bind to or modulate the activity of CGA7 or CJA8 and evaluate the effect of a candidate colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit the activity of CGA7 or CJA8, respectively, and is used to screen for an agent that can interfere with the binding of CGA7 or CJA8 to the antibody. The antibody can be used to treat colorectal cancer. The antibody or a fragment of it is used to localize a therapeutic group to a colorectal cancer tissue, where the therapeutic group is a cytotoxic agent or a radioisotope. Antisense molecules are used to inhibit colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or CJA8 are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it are used to elicit an immune response. CGA7 or CJA8 is used to determine the prognosis of an individual with colorectal cancer. Nucleic acid encoding CGA7 or CJA8 can be used in vaccines

SQ Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 U; 0 Other;

Query Match 99.0%; Score 1292; DB 6; Length 2079;

Best Local Similarity 99.9%; Pred: No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy	3	GGATCTGTGACGTGATCAAACTCTTGAACAGCCTCGATGTCAAAACCCCTGTGGGCAAAACCCGG	52
Db	216	GGATCTGTGACGTGATCAAACTCTTGAACAGCCTCGATGTCAAAACCCCTGTGGCAAAACCCGG	275
Qy	63	TATCCCCATGAGAGACCTTCAGAAAAG-GTGGGGATCCCCATCATCATAGCACTACTTGAGCC	121
Db	276	TATCCCCATGAGAGACCTTCAGAAAAGTGGGGATCCCCATCATCATAGCACTACTTGAGCC	335
Qy	122	TGGCAGATATCATATGTGTGTCTCATCAAGTGATTTCTGGATTAATAACTACTTCTCC	181
Db	336	TGGCAGATATCATATGTGTGTCTCTCATCAAGGTGATTTCTGGATTAATAACTACTTCTCC	395
Qy	182	TCTCGGGGAGCCTCTCCACTTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGACT	241
Db	396	TCTCGGGGAGCCTCTCCACTTTCATCCCGAGGAAGCAGCTGTGTGACGAGAGCTGGACT	455
Qy	242	GTCCCTTGGGGAGGACGAGGAGCACTGTGTCTAAGAGCTTCCCGGAAGGGCTGTGAGTGG	301
Db	456	GTCCCTTGGGGAGGACGAGGAGCACTGTGTCTAAGAGCTTCCCGGAAGGGCTGTGAGTGG	515
Qy	302	CAGTCCGGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGGAACT	361
Db	516	CAGTCCGGCTCTCCAAAGGACCGATCCACACTGCAGGTGTGTGACTCGGCCACAGGGAACT	575
Qy	362	GGTTCTCTGCCTGTTTGGAACTTCCAGAAAGCTCTCGCTGAGACAGCCTGTAGGCAGA	421
Db	576	GGTTCTCTGCCTGTTTGGAACTTCCAGAAAGCTCTCGCTGAGACAGCCTGTAGGCAGA	635
Qy	422	TGGGCTACAGCAGAGAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTGG	481
Db	636	TGGGCTACAGCAGAGAAACCCACTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTGG	695
Qy	482	ATGTTGTTGAAATCACAGAAAAACGACGAGGCTTCGCATGCGGAACCTCAAGTGGGGCCT	541
Db	696	ATGTTGTTGAAATCACAGAAAAACGACGAGGCTTCGCATGCGGAACCTCAAGTGGGGCCT	755
Qy	542	GTCTCTCAGGCTCCCTGGTCTCCCTGCACGTCTTCCCTGTGGGAGAGCCTTGAGAGCC	601
Db	756	GTCTCTCAGGCTCCCTGGTCTCCCTGCACGTCTTCCCTGTGGGAGAGCCTTGAGAGCC	815
Qy	602	CCCTGTGGTGGGTGGGAGGAGCCCTCTCTGGATTCTTTGGCTTTGGCAGGTCAGCATCC	661
Db	816	CCCTGTGGTGGGTGGGAGGAGCCCTCTCTGGATTCTTTGGCTTTGGCAGGTCAGCATCC	875

28-JAN-2000; 2000US-00493444.
15-MAR-2000; 2000US-00525993.
(EOSB-) EOS BIOTECHNOLOGY INC.
Mack D, Gish KC, Wilson KE;
WPI; 2003-066245/06.
P-PSDB; ABG72428.
Diagnosing colorectal cancer comprises determining the expression of a
gene encoding CUA8 in a first colon tissue of a first individual, and
comparing them to the expression of a gene encoding CUA8 in a second
normal tissue.
Claim 11; Fig 1; 31pp; English.
The invention describes a method of diagnosing colorectal cancer
comprising determining the expression of a gene encoding CUA8 or its
fragment in a first colon tissue of an individual, and comparing the
expression of the gene to the expression of the gene in normal tissue
from the individual or in tissue from a second unaffected individual. A
difference in the expression indicates that the first individual has
colorectal cancer. The CUA8 genes and proteins are useful for diagnosing
and prognosticating colorectal cancer. The methods are also useful for
screening candidate bioactive agents that can modulate colorectal cancer,
or for treating or inhibiting colorectal cancer in a patient e.g. by gene
therapy. This sequence encodes the human colorectal cancer modulating
protein CUA8
Sequence 2079 BP; 489 A; 594 C; 575 G; 421 T; 0 U; 0 Other;
Query Match 99.0%; Score 1292; DB 7; Length 2079;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 3 GGATCCTCAGAGTATCAACCTCTGAACAGAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 62
DB 216 GGATCCTGACAGTATCAACCTCTGAACAGAGCTCGATGTCAAAACCCCTGCGCAAAACCCCG 275
QY 63 TATCCCATGGAGACCTTCAGAAAG- GTGGGATGCCCATCATATAGCACTACTAGACC 121
DB 276 TATCCCATGGAGACCTTCAGAAAGTGTGGGATGCCCATCATATAGCACTACTAGACC 335
QY 122 TGGCGAGTATCATATTGGTGTGCTCATCAAGTGTATCTGTGATAAATACTACTTCC 181
DB 336 TGGCGAGTATCATATTGGTGTGCTCATCAAGTGTATCTGTGATAAATACTACTTCC 395
QY 182 TCTGGGCGAGCCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGA CGGAGAGCTGGACT 241
DB 396 TCTGGGCGAGCCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGA CGGAGAGCTGGACT 455
QY 242 GTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTGG 301
DB 456 GTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTGCAGTGG 515
QY 302 CAGTCCGCTCTCCAGGAGCCGATCCACTGAGGTGTGGACTCGGCCACAGGGAAC 361
DB 516 CAGTCCGCTCTCCAGGAGCCGATCCACTGAGGTGTGGACTCGGCCACAGGGAAC 575
QY 362 GGTCTCTGCCCTGTTTTCGACAACTTCAAGAGCTCTGCTGAGACAGCCTGTAGGCAGA 421
DB 576 GGTCTCTGCCCTGTTTTCGACAACTTCAAGAGCTCTGCTGAGACAGCCTGTAGGCAGA 635
QY 422 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGATCTGG 481
DB 636 TGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCACAGGATCTGG 695
QY 482 ATGTTTGTGAAATTCAGAAAAACAGCCAGGAGCTTGGCATGCGGAACCAAGTGGGCCCT 541
DB 696 ATGTTTGTGAAATTCAGAAAAACAGCCAGGAGCTTGGCATGCGGAACCAAGTGGGCCCT 755
QY 542 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGGCTGTGGGAGAGCCCTGAAGACCC 601

DB 756 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGGCTGTGGGAGAGCCCTGAAGACCC 815
QY 602 CCGTGTGTGGGTGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTGAGCATCC 661
DB 816 CCGTGTGTGGGTGGGAGGAGGCTCTGTGTGATTTCTTGGCTTGGCAGGTGAGCATCC 875
QY 662 AGTACGACAAACAGCAGCTCTGTGTGAGGAGGAGCATCTGTGACCCCACTGGTCCCTCACCG 721
DB 876 AGTACGACAAACAGCAGCTCTGTGTGAGGAGGAGCATCTGTGACCCCACTGGTCCCTCACCG 935
QY 722 CAGCCCACTGCTTCAGAAAAACATACCGATGTGTCAACTTGAAGGTGCGGGAGGCTCAG 781
DB 936 CAGCCCACTGCTTCAGAAAAACATACCGATGTGTCAACTTGAAGGTGCGGGAGGCTCAG 995
QY 782 ACAAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATGATTAATCAACC 841
DB 996 ACAAACCTGGGAGCTTCCCATCCCTGGCTGTGGCCAGATCATCATGATTAATCAACC 1055
QY 842 CCATGTACCCCAAGAGCAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCCTTTCT 901
DB 1056 CCATGTACCCCAAGAGCAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCACTCCTTTCT 1115
QY 902 CAGGCAAGTTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 961
DB 1116 CAGGCAAGTTCAGGCCCATCTGTCTGCCCTTCTTGTATGAGGAGCTCACTCCAGCCACCC 1175
QY 962 CACTCTGGATCATTCGATGGGCTTTACGAAGCAGATGAGGGAGAGATGCTGTGACATAC 1021
DB 1176 CACTCTGGATCATTCGATGGGCTTTACGAAGCAGATGAGGGAGAGATGCTGTGACATAC 1235
QY 1022 TGCTGCAAGGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGACAGAGATGCGTACC 1081
DB 1236 TGCTGCAAGGCTCAGTCCAGGTCAATTGACAGCACACGGTGCATATGACAGAGATGCGTACC 1295
QY 1082 AGGGGAGATCACCGAGAGATGATGTGTGAGGCAATCCCGAAGGGGTGTGAGACCT 1141
DB 1296 AGGGGAGATCACCGAGAGATGATGTGTGAGGCAATCCCGAAGGGGTGTGAGACCT 1355
QY 1142 GCCAGGTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCA 1201
DB 1356 GCCAGGTGACAGTGTGGGCCCTCTGATGTACCAATCTGACCAAGTGGCATGTGTGGGCA 1415
QY 1202 TCGTTAGCTGGGCTATGGCTGGGGGGCCGAGGACCCCGAGGAGTATACCAAGGTCT 1261
DB 1416 TCGTTAGCTGGGCTATGGCTGGGGGGCCGAGGACCCCGAGGAGTATACCAAGGTCT 1475
QY 1262 CAGCCTATCTCAATCGATCTACAATGTCTGGAAGGCTGAGCTG 1305
DB 1476 CAGCCTATCTCAATCGATCTACAATGTCTGGAAGGCTGAGCTG 1519

Search completed: June 7, 2004, 23:47:00
Job time : 579 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 23:21:59 ; Search time 3766 Seconds
(without alignments)
10347.888 Million cell updates/sec

Title: US-10-030-688-1

Perfect score: 1305

Sequence: 1 atggatcctgacagtgatca.....atgtctggaagctgagctg 1305

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_nam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rtd.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1009	77.3	1973	11	BC004855
2	993.4	76.1	995	29	AV420489
3	835.6	64.0	995	29	AV420490
4	805.8	61.7	2177	11	AK078890

ALIGNMENTS

RESULT 1
BC004855
LOCUS Homo sapiens, Similar to transmembrane protease, serine 4, clone
DEFINITION BC004855.1 173 bp mRNA linear HTC 12-JUL-2001
ACCESSION BC004855
VERSION BC004855.1 GI:14709533
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1973)
AUTHORS Strausberg R.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgaps@remail.nih.gov
Tissue Procurement: DCID/STP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>

5 716.6 54.9 995 29 AY420491
6 699.8 53.6 844 10 BE531100
7 634.2 48.6 975 12 EG288427
8 598.4 45.9 895 13 BU522841
9 587.8 45.0 617 12 BM795149
10 584.6 44.8 920 10 BE616663
11 551.2 42.2 854 12 EG822314
12 538.2 41.2 569 12 BM686292
13 537.2 41.2 1007 12 EG386903
14 528.2 40.5 956 13 BU523218
15 513.6 39.4 787 10 BE615750
16 502.8 38.5 879 10 BE616186
17 495 37.9 548 14 CB159713
18 484.2 37.9 503 10 BE514663
19 488.8 37.5 742 12 EG468475
20 481.4 36.9 498 12 BM795128
21 479.8 36.8 744 12 BM008802
22 479.4 36.7 515 14 CB142902
23 474.8 36.4 935 10 BE615106
24 474.6 35.1 920 13 BQ922656
25 456.2 35.0 931 13 BU157248
26 453.8 34.8 977 10 BE614660
27 442.2 33.9 689 12 EG966811
28 440.8 33.8 659 9 AI924527
29 439.4 33.7 648 9 AI924182
30 437.2 33.5 645 10 BF118991
31 436.4 33.4 1105 10 BE736425
32 433.2 33.2 482 12 EG984172
33 428.4 32.8 803 14 CA311909
34 419.8 32.2 640 12 EG822651
35 406 31.1 980 10 BE745289
36 383.4 29.4 936 12 BG387113
37 381.8 29.3 1088 10 BE616293
38 373.4 28.6 408 12 BM762292
39 368.8 28.3 435 10 BF811390
40 360.2 27.6 850 12 B1251465
41 355 27.2 584 9 AI597567
42 353.8 27.1 693 12 BG481239
43 353 27.0 711 10 BE937482
44 352 27.0 644 12 BM791640
45 344 26.4 721 13 BU600097

BC004855. 173 bp mRNA linear HTC 12-JUL-2001
Homo sapiens, Similar to transmembrane protease, serine 4, clone
IMAGE:3835263, mRNA.

BC004855.1 GI:14709533

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 1973)

Strausberg R.

Direct Submission

Submitted (21-MAR-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps@remail.nih.gov

Tissue Procurement: DCID/STP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbiology.org>

contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: p Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8347148
This clone has the following problem: frame shifted.

FEATURES

source

1. 1973
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3835263"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH MGC_9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Query Match 77.3%; Score 1009; DB 11; Length 1973;
Best Local Similarity 89.0%; Pred. No. 1.4e-236;
Matches 1162; Conservative 0; Mismatches 0; Indels 143; Gaps 1;

Qy 1 ATGGATCCTGACAGTGATCAACCTCTGAACAGGCTCGATGTCAAACCCCTGGCGAAACCC 60
Db |
Qy 243 ATGGATCCTGACAGTGATCAACCTCTGAACAGGCTCGATGTCAAACCCCTGGCGAAACCC 302
Db |
Qy 61 CGTATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATATAGCACTACTTGAGC 120
Db |
Qy 303 CGTATCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATATAGCACTACTTGAGC 362
Db |
Qy 121 CTGGGAGTATCATATGTTGTTCTCATCAAGTGAATCTGGATAAATACTACTTC 180
Db |
Qy 363 CTGGGAGTATCATATGTTGTTCTCATCAAGTGAATCTGGATAAATACTACTTC 422
Db |
Qy 181 CTCTGGGCGAGCTCTCCACATTCATCCGAGGAGCAGCTGTGACGAGAGCTGGAC 240
Db |
Qy 423 CTCTGGGCGAGCTCTCCACATTCATCCGAGGAGCAGCTGTGACGAGAGCTGGAC 482
Db |
Qy 241 TGTCCCTTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCCAGTG 300
Db |
Qy 483 TGTCCCTTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTCCAGTG 542
Db |
Qy 301 GCGATCGGCTCTCCAGAGCAGTCCACACTGCAGGCTCGGACTCGGCGACAGGAGAC 360
Db |
Qy 543 GCGATCGGCTCTCCAGAGCAGTCCACACTGCAGGCTCGGACTCGGCGACAGGAGAC 602
Db |
Qy 361 TGTCTCTGCTGCTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAG 420
Db |
Qy 603 TGTCTCTGCTGCTTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAG 662
Db |
Qy 421 ATGGGCTACAGCGCAAAACCCACTTCAGAGCTGTGGAGATGGGCGCCAGACAGAGATCTG 480
Db |
Qy 663 ATGGGCTACAGCGCAAAACCCACTTCAGAGCTGTGGAGATGGGCGCCAGACAGAGATCTG 722
Db |
Qy 481 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGCATCCGAACTCAAGTGGGCCC 540
Db |
Qy 723 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGCATCCGAACTCAAGTGGGCCC 782
Db |
Qy 541 TGTCTCTAGGCTCCCTGTCTCCCTGCACTGTGCTGTGGAGAGAGCTGGAAGACC 600
Db |
Qy 783 TGTCTCTAGGCTCCCTGTCTCCCTGCACTGTGCTGTGGAGAGAGCTGGAAGACC 842
Db |
Qy 601 CCCCGTGTGGTGGGAGAGGAGGCTCTGTGGATTTCTTGCCCTTGGAGCTCAGCATC 660
Db |
Qy 843 CCCCGTGTGGTGGGAGAGGAGGCTCTGTGGATTTCTTGCCCTTGGAGCTCAGCATC 902
Db |
Qy 661 CAGTAGCAAAACAGCACCTCTGTGGAGGAGCATCTTGAGACCCCTGAGGCTCTCAG 720
Db |
Qy 903 CAGTAGCAAAACAGCACCTCTGTGGAGGAGCATCTTGAGACCCCTGAGGCTCTCAG 962
Db |

Qy 721 GCAGCCCACTGCTTTCAGAAACATACCATGATGTGTCACTGGAGGTGGCGGAGGCTCA 780
Db |
Qy 963 GCAGCCCACTGCTTTCAGAAACATACCATGATGTGTCACTGGAGGTGGCGGAGGCTCA 1022
Db |
Qy 781 GACAACTGGGAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATATGATTAATCAAC 840
Db |
Qy 1023 GACAACTGGGAGCTTCCCATCCCTGGCTGGCCCAAGATCATCATATGATTAATCAAC 1082
Db |
Qy 841 CCCATGTACCCCAAGAGCAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCTCCTTTC 900
Db |
Qy 1083 CCCATGTACCCCAAGAGCAATGACATCGCCCTCATGAAGCTGCAGTTCCTCCTCCTTTC 1142
Db |
Qy 901 TCAGGCAAGTTCAGGCGCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCATCCAGGCACC 960
Db |
Qy 1143 TCAGGCAAGTTCAGGCGCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCATCCAGGCACC 1202
Db |
Qy 961 CCATCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAGATGTCTGACATA 1020
Db |
Qy 1203 CCATCTGGATCATTTGATGGGCTTTACGAAGCAGAAATGGAGGAGATGTCTGACATA 1243
Db |
Qy 1021 CTGTGCGAGCGTCAAGTCCAGGTCATTTGACGACACAGGTCGAATGCAGACGATGCGTAC 1080
Db |
Qy 1244 ----- 1243
Qy 1081 CAGGGGAAGTCCAGGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACC 1140
Db |
Qy 1244 ----- 1243
Qy 1141 TGCCAGGCTGACAGTGTGGGCCCCCTGATGATCAATCTCACCAGTGGCATGTGGTGGGC 1200
Db |
Qy 1244 ----AGGCTGACAGTGTGGGCCCCCTGATGATCAATCTCACCAGTGGCATGTGGTGGGC 1299
Db |
Qy 1201 ATGTTAGCTGGGCTATGCTGGGGGCCCCAGAGACCCAGGAGTATACACAGGCTC 1260
Db |
Qy 1300 ATGTTAGCTGGGCTATGCTGGGGGCCCCAGAGACCCAGGAGTATACACAGGCTC 1359
Db |
Qy 1261 TCAGCTATCTCAACTGGATCTCAATGCTTGAAGGCTGAGCTG 1305
Db |
Qy 1360 TCAGCTATCTCAACTGGATCTCAATGCTTGAAGGCTGAGCTG 1404
Db |

RESULT 2

AY420489

LOCUS

DEFINITION

Homo sapiens

Genomic survey sequence.

AY420489

VERSION

AY420489.1

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 995)

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,

Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,

Adams, M.D. and Cargill, M.

Submitted (16-NOV-2003)

Rockville, MD 20850, USA

This sequence is made by sequencing genomic exons and ordering them

based on alignment.

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

COMMENT

[illegible]

Db 241 NNNNTCTCGCTGAGACAGCCCTGTAGGACAGATGGCTACAGCAGCAACCCACTTTTCAGAG 300
Qy 452 CTGTGAGATGGCCAGACAGCAGATCTGATGTTGTTGAATACAGAAAACAGCCAGG 511
Db 301 CTGTGAGATGGCCAGACAGCAGATCTGATGTTGTTGAATACAGAAAACAGCCAGG 360
Qy 512 AGCTTCGATCGCGAATCAAGTGGCCCTGTCTCTCAGGCTCCCTGCTGCTCCCTGCAC 571
Db 361 AGCTTCACATCGCGAATCAAGTGGCCCTGTCTCTCAGGCTCCCTGCTGCTCCCTGCAC 420
Qy 572 GTCTTCGCTGTGGAGAGAGCCTGAAGACCCCTGTGTGTGGTGGGAGGAGGCTCTGTG 631
Db 421 GTCTTCGCTGTGGGAGAGCCTGANNNCCCTGTGTGTGGTGGGAGGAGGCTCTGTG 480
Qy 632 TGGATCTTGGCTTGGCAGCTCAGATCCAGTACGACAAACAGCAGCAGTCTGTGGAGGGA 691
Db 481 TGGATCTTGGCTTGGCAGCTCAGATCCAGTACGACAAACAGCAGCAGTCTGTGGAGGGA 540
Qy 692 GCATCTCGACCCCACTGGCTCTCACGGCAGCCCACTGCTTCAGGAAACATACCGATG 751
Db 541 GCATCTCGACCCCACTGGCTCTCACGGCAGCCCACTGCTTCAGGAAACATACCGATG 600
Qy 752 TGTTCAGTGGAGGTGGCGGAGGCTCAGACAACTGGCGAGCTCCCATCCCTGGCTG 811
Db 601 TGTTCAGTGGAGGTGGCGGAGGCTCAGACAACTGGCGAGCTCCCATCCCTGGCTG 660
Qy 812 TGGCCAGATCATCATCATGTAATCAACCCCACTGATCCCAAGCAATGATGATCGCC 871
Db 661 TGGCCAGATCATCATCATGTAATCAACCCCACTGATCCCAAGCAATGATGATCGCC 720
Qy 872 TATGAGCTGAGTTCCTCCACTCACTTCTCAGGACAGTCAAGCCCATCTCTGCTGCT 931
Db 721 TATGAGCTGAGTTCCTCCACTCACTTCTCAGGACAGTCAAGCCCATCTCTGCTGCT 780
Qy 932 TCTTCATGAGGAGTCACTTCAGCCACCCCACTCTGGATCATTTGATGGGCTTTTACGA 991
Db 781 TCTTCATGAGGAGTCACTTCAGCCACCCCACTCTGGATCATTTGATGGGCTTTTACGA 840
Qy 992 AGCAGATCGGAGGAGATGTCTGATCATCTCTCAGGAGTCAAGCCCATCTGATGACA 1051
Db 841 AGCAGATCGGAGGAGATGTCTGATCATCTCTCAGGAGTCAAGCCCATCTGATGACA 900
Qy 1052 GCACAGCTGCAATGAGACAGTGGTACAGGGGAGTCAAGCCCATCTGATGATGCTGTG 1111
Db 901 GCACAGCTGCAATGAGACAGTGGTACAGGGGAGTCAAGCCCATCTGATGATGCTGTG 960
Qy 1112 CAGGCATCCCGAAGGGGTGTGGACACCTGCGAG 1146
Db 961 CAGGCATCCCGAAGGGGTGTGGACACCTGCGAG 995

RESULT 4
AK078890
LOCUS
DEFINITION
Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:903062G02 product:SIMILAR TO TRANSMEMBRANE PROTEASE, SERINE 4 homolog [Mus musculus], full insert sequence.
AK078890
VERSION
AK078890.1 GI:26098158
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kenno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2177)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Takawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers
1. 2177
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:903062G02"
/db_xref="MGI:2397607"
/db_xref="taxon:10090"
/clone="903062G02"
/sex="male"
/tissue_type="colon"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
misc_feature
304..1507
/notes="SIMILAR TO TRANSMEMBRANE PROTEASE, SERINE 4 homolog

COMMENT

FEATURES
source

[Mus musculus] (SPTF|AAH21368, evidence: FASTY, 92%ID, 100%length, match=1201)
putative
2154..2159
/note="putative"
2177
/note="putative"

polyA_signal
polyA_site

ORIGIN

Query Match 61.7%; Score 805.8; DB 11; Length 2177;
Best Local Similarity 82.7%; Pred. No. 1.3e-186;
Matches 921; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 193 CCTCTCCACTTCATCCCGAGGAGCAGCTGTGTGACGAGAGCTGTGCTCTTGGG 252
Db 392 CCCCTGACCTTCATTCAGAGGGGCCAGTTGTGTGACGGCCACCTTACTGCGCTCAGG 451

QY 253 GAGGACGAGGACCTGTGTCAAGAGCTTCCCGAAGGCGCTGCAGTGCAGCTCGCTC 312
Db 452 GAGGATGAGGAACACTGTGTCAAGAGCTTCCCTGAAAGCCCGAGTGGCAGTCCGGCTC 511

QY 313 TCCAGGACCGATCCACACTGCAGGTGTGACTCGGCCACAGGAACTGGTTCTCTGCC 372
Db 512 TCCAGGACAGATCCACCTGCAGGTGTGATGACGACACAGGACCTGGGCTCAGTC 571

QY 373 TGTTCGACACTTCACAGAGCTTCGTGTGACAGAGCTGTAGGACAGTGGCTACAGC 432
Db 572 TGTTCGACACTTCACAGAGCTTCGTGTGACAGAGCTTCGACAGAGTGGCTATGAC 631

QY 433 AGCAAAACCACTTCAGAGCTGTGAGATGTCGACAGGATCTGGATGTGTGAA 492
Db 632 ACCAGCCGCTTCAGAGCAGTGGAGATCCGTCCAGATCAGAACCTCCCTGTGCTCAA 691

QY 493 ATCAGAGAAACAGCCAGGAGCTTCGATGCGGAATCAAGTGGGCGCTGTCTCAGGC 552
Db 692 GTCACAGGAAACAGCCAGGAACTTCAGGTGCAAGATGGAAGCAGATCTGCTCTCAGGC 751

QY 553 TCCCTGCTCCCTGCACCTGTCTCCCTGTGGAGAGGCTGAGACCCCGCTGTGGTG 612
Db 752 TCCCTGCTTCCTGCGCTGCTTCACTGTGAAAGAGCTTGAAGACCTCTCGTGTGGTG 811

QY 613 GGTGGGAGGAGGCTCTGTGATCTTGGCTTGGCAGGTTCAGATCAGTACAGCAAA 672
Db 812 GGTGGGAGGAGGCTCTGTGATCTTGGCTTGGCAGGTTCAGATCAGTACAGCAAG 871

QY 673 CAGCAGCTGTGTGAGGAGAGCTTCGTGACCCCTGCTGCTCAGGAGCCCTACGTC 732
Db 872 CAGCAGCTGTGTGAGGAGAGCTTCGTGACCCCTGCTGCTCAGGAGCCCTACGTC 931

QY 733 TTCAGAAACATACCGATGTGTTCACCTGGAAGGTGCGGCGAGGCTCAGACAACTGGGC 792
Db 932 TTCAGAAAGTATCTTGATGTGTCAAGCTGGAAGGTTCAGGCGAGGCTCAACATCTGGGT 991

QY 793 AGCTTCCCATCCCTGGCTGTGGCCAGATCATCATCATTTGAATTCACCCCATGTACCCC 852
Db 992 AACTCTCCATCTCTGCTGTGGCCAGATCTTCATCGCTGAACCAATCTCTGTACCCC 1051

QY 853 AAAGCAATGACATCGCCCTCATGAGCTGCAAGTTCCTCCACTCAGTCTTCAGGACAGTC 912
Db 1052 AAAGCAAGGACATTCCTCTTGAAGCTGCAAGTTCCTCCACTCAGTCTTCAGGCTCAGTC 1111

QY 913 AGGCCCATCTGTCTGCTCTTCTTGTGAGGAGCTCACTTCCAGCCACCCCACTCTGGATC 972
Db 1112 AGGCCCATCTGTCTGCTCTTCTTGTGAGGAGCTCACTTCCAGCCACCCCACTCTGGATC 1171

QY 973 ATTGATGGGCTTTACGAGAGGATGAGGAGAGATGTCTGATATCTCTGCGAGGG 1032
Db 1172 ATTGATGGGCTTTACAGAGAAACCGAGAGAGATGTCTGATGATCTCTGCGAGGA 1231

QY 1033 TCAGTCCAGCTCATTGACAGCAGCAGGTGCAATGACAGAGTGCCTTACAGGGGAGATC 1092
Db 1232 TCAGTCCAGCTCATTGACAGCAGCAGGTGCAATGACAGAGTGCCTTACAGGGGAGATG 1291

QY 1093 ACCGAGAGATGATGTGTGAGGATCCCGAAGGGGTGTGGACACCTGCGAGGTGAC 1152
Db 1292 ACCGCTGAGATGCTGTGTGAGGATCCCGAAGGGGTGTGGACACCTGCGAGGTGAC 1351

QY 1153 ASTGTTGGGCTCTGATGTACCAATCTGACCATGTGTGTGGGCTGTGTGTAGCTGG 1212
Db 1352 AGTGGTGGGCTTGTGTGTACCAATCTGACCAATGTGAGGTGTGTGTAGCTGG 1411

QY 1213 GGTATGGCTGCGGGGCGGACGACCCAGGAGTATACACCAAGGTCTCAGCTATCTC 1272
Db 1412 GGCATGGATGCGGGCGGCAAGTACTCTGTGAGTGTATACCAAGGTCACTGCTATCTC 1471

QY 1273 AACTGATCTACCAATCTCTGGAAGGTGTGAGCTG 1305
Db 1472 AACTGATCTACCAATCTCTGGAAGGTGTGAGATG 1504

RESULT 5
AY420491 995 bp DNA linear GSS 17-DEC-2003
LOCUS Mus musculus TWPRSS4 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY420491
VERSION AY420491.1 GI:39776448
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 995)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 995)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source Location/Qualifiers
1..995
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..5995
/gene="TWPRSS4"
/locus_tag="HGM7241"

gene
ORIGIN

Query Match 54.9%; Score 716.6; DB 29; Length 995;
Best Local Similarity 82.5%; Pred. No. 6.9e-165;
Matches 821; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 152 TCAAGGTGATTCGATTAATACTACTTCTCTGCGGCGAGCTCTCCACTTCATCCCGA 211
Db 1 TCAAGGTGATTCGATTAATACTACTTCTCTGCGGCGAGCTCTCCACTTCATTCAGA 60

QY 212 GGAAGCAGCTGTGTGACGAGAGCTGACCTGCTCCCTGGGGAGGAGGAGGAGCAGCTGTG 271
Db 61 GGGGCCAGTGTGTGTGACGGCCACCTTGTACCTGCGCTCAGGGGAGGATGAGGAACACTGTG 120

QY 272 TCAAGAGCTTCCCGAAGGGCCTGCACTGGCAGTCCCGCTCTCCAGGACCGGATCCACAC 331
Db 121 TCAAGAGCTTCCCGAAGGGCCTGCACTGGCAGTCCCGCTCTCCAGGACGATCCACCC 180

332 TGCAAGTGTGGAGTCCGGCCACAGGAACTGTCTCTGCTGTGTTTGCACAACTTTCACAG 391
 Db TGCAGGTGTGGATGATGAGCCACAGAGGACCTGGGCTCAGTGTGTTTGCACAACTTTCACAG 240
 392 AGCTCTCCCTGAGACAGCTCTAGGACAGATGGGTACAGGACGAAACCCACTTTCACAG 451
 Db AAGACTTGGCCAAAGACAGCTCTAGGACAGATGGGTACAGGACGAAACCCACTTTCACAG 300
 452 CTGTGAGATTTGGCCACAGACAGGATCTGGATGTTGTTGAAATACAGAAACACCCAGG 511
 Db CAGTGGAGATCCGTCCAGATCAGAACCTCCCTGTTGCTCAAGTACAGGAAACACCCAGG 360
 512 AGCTTCGATGGGAAGTCAAGTGGGCTCTGTCTCAGGCTCCCTGTCTCCCTGCACT 571
 Db AACTTCAGGTGACAGATGGAAGCAGATCCTGCTCTCAGGCTCCCTGTCTCCCTGCACT 420
 572 GTCTTCCTGTGGGAAGAGCTGAACACCCCTGCTGTGGTGGGAGGAGGCTCTG 631
 Db GCCTTCAGTGTGGAAGAGCTGAAGACTCTCTGTGTGGTGGGAGGAGGCTCTG 480
 632 TGAATCTTGGCTTGGCAGGTGAGATCAGATCAGTACAGCAAAACAGACAGTCTGTGGAGGA 691
 Db TGAATCTTGGCTTGGCAGGTGAGATCAGATCAGTACAGCAAAACAGACAGTCTGTGGAGGA 540
 692 GCATCTGGACCCCACTGGTCTCAGCGGACCCACTGCTTACAGGAAACATACCGATG 751
 Db GCATCTGGATCCCACTGGATCTCAGAGAGCCCACTGCTTACAGGAAATCTTGATG 600
 752 TGTTCAGTGGAGGTGCGGAGGCTCAGACAACTGGGAGCTTCCCATCCCTGGCTG 811
 Db TGTTCAGTGGAGGTGCGGAGGCTCAGACAACTGGGAGCTTCCCATCCCTGGCTG 660
 812 TGGCCAGATCATCATATGATTCACCCCATGATGACCCCAAGACATGACATCGCCC 871
 Db TGGCCAGATCATCATATGATTCACCCCATGATGACCCCAAGACATGACATCGCCC 720
 872 TCATGAAGCTGCAGTTCCTCCACTCAGTTCAGGACAGTCAAGGCTTCTGTCTGCTGCT 931
 Db TTGTTAAGCTGCAGTTCCTCCACTCAGTTCAGGACAGTCAAGGCTTCTGTCTGCTGCT 780
 932 TCTTTGATGAGAGTCACTCAGCCACCCCACTCTGATCATTTGATGGGCTTTACGA 991
 Db TCTCTGATGAGTGTCTGTCAGGACACCACTCTGCTGATGATGGGCTTTACGA 840
 992 AGCAGAAATGGAGGAAGTGTCTGACATCTGCTGAGGACAGTCAAGGCTTCTGTGCTGACA 1051
 Db AAGAAACGGAGGAAGATGTCTGATCTGCTGATCTGATCTGATCTGATCTGATCTGAT 900
 1052 GCACAGGTGCAATGAGACAGTGGGTACCGGGGGAAGTCAAGGAGGATGATGTTG 1111
 Db GCACAGGTGCAATGAGACAGTGGGTACCGGGGGAAGTCAAGGAGGATGATGTTG 960
 1112 CAGGCATCCCGAGGGGTGGACACCTGCCAG 1146
 Db CAGGTACCCACAGGTTGGCAAGGACCTGCCAG 995
 RESULT 6
 BE531100
 LOCUS 844 bp mRNA linear EST 09-AUG-2000
 DEFINITION 601278466F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610337 5',
 mRNA sequence.
 ACCESSION BE531100
 VERSION BE531100.1 GI:9759745
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 844)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-x@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM266 row: d column: 18
 High quality sequence stop: 712.
 Location/Qualifiers
 1..844
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3610337"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_39"
 /notes="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;
 Site: 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies)."
 ORIGIN
 Query Match 53.6%; Score 699.8; DB 10; Length 844;
 Best Local Similarity 96.4%; Pred. No. 84e-161;
 Matches 749; Conservative 0; Mismatches 22; Indels 6; Gaps 3;
 118 AGCTGCGAGTATCATATGTTGGTGTCTCTCATCAAGGTGATTCGTAATAACTAC 177
 Db 1 AGCTGCGAGTATCATATGTTGGTGTCTCTCATCAAGGTGATTCGTAATAACTAC 60
 178 TTCTCTGCGGAGCTCTCCACTTCATCCGAGGACGAGCTGTGTGACGAGAGCTG 237
 Db 1 TTCTCTGCGGAGCTCTCCACTTCATCCGAGGACGAGCTGTGTGACGAGAGCTG 120
 238 GACTGTCTCTGCGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCTGCA 297
 Db 1 GACTGTCTCTGCGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGCTGCA 180
 298 GTGCGAGTCCGCTCTCCAGGACCGATCCACACTGCGAGTCTGCGCCACAGG 357
 Db 1 GTGCGAGTCCGCTCTCCAGGACCGATCCACACTGCGAGTCTGCGCCACAGG 240
 358 AACTGTGTTCTCTGCTGTTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCTTAG 417
 Db 1 AACTGTGTTCTCTGCTGTTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCTTAG 300
 418 CAGATGGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGTTGGCCACAGCAGAT 477
 Db 1 CAGATGGGCTACAGCAGCAACCCACTTTCAGAGCTGTGGAGTTGGCCACAGCAGAT 360
 478 CTGATGTTGTTGAATTCACAGAAACACAGCAGAGCTTCGATGCGGAACTCAAGTGG 537
 Db 1 CTGATGTTGTTGAATTCACAGAAACACAGCAGAGCTTCGATGCGGAACTCAAGTGG 420
 538 CCTGTCTCTCAGGCTCCCTGTGCTCCCTGCACTGTCTGCTGCTGGAGAGGCTGAAG 597
 Db 1 CCTGTCTCTCAGGCTCCCTGTGCTCCCTGCACTGTCTGCTGCTGGAGAGGCTGAAG 480
 598 ACCCCCGCTGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTTGGCAGGTGAG 657
 Db 1 ACCCCCGCTGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCTTTGGCAGGTGAG 540
 658 ATCCAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGAGACCCCACTGGGCTCTC 717
 Db 1 ATCCAGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGAGACCCCACTGGGCTCTC

```

Db      541 ATCCAGTAGCAGCAACAGCAGCTCTGTGGAGGGAGCATCTTGGACCCCACTGGGTCTCTC 600
Qy      718 ACGGAGGCCACTGTCTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGCAGGC 777
Db      601 ACGGAG-CCACTGCTTCAGGAAACATACCGATGTGTTCAACTGGAAGGTGCGGCGCAGGC 659
Qy      778 TCAGCAAACTGGGAGCTCCCTCCCTGGCTGTGGCCAGATCATCATATTGAATTC 837
Db      660 TCAGAC-AACTGGGAGCTTCCCATCCCTGGGCTGTGGCCAGATCATCATATTGAATTC 718
Qy      838 AACCCCATGTACCCCAAGACAAATGACATCGGCCCTCATGAAGCTGCAGTTCCTCACTC 894
Db      719 AACC-----TGTTCCTCCAGACATGACATCGCCCTCATGAAGCTGGTTCCTCCCTCAC 771

RESULT 7
BG288427
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Bg288427 975 bp mRNA linear EST 21-FEB-2001
60238091F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4516990 5',
mRNA sequence.
Bg288427
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 975)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10409 row: e column: 23
High quality sequence stop: 659.
Location/Qualifiers
1..975
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4516990"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 93"
/note="Organ: bladder; Vector: pCMV-SPORT6; Site 1: Not I;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 48.6%; Score 634.2; DB 12; Length 975;
Best Local Similarity 96.4%; Pred. No. 1.2e-144;
Matches 560; Conservative 0; Mismatches 23; Indels 2; Gaps 1;

Qy      584 GGAAGAGCTGGAAGACCCCTGTGGTGGTGGGAGAGCCCTCTGTGGATCTTTGGC 643
Db      1 GGAAGAGCTGGAAGACCCCTGTGGTGGTGGGAGAGGCCCTCTGTGGATCTTTGGC 60
Qy      644 CTTCGAGCTCAGATCCAGTACGACAAACACACAGCTGTGTGGAGGAGCATCTCGGACC 703
Db      61 CTTCGAGCTCAGATCCAGTACGACAAACACAGCTGTGTGGAGGAGCATCTCGGACC 120
Qy      704 CCACCTGGGTCTCAGCGAGCCCACTGTTTCAGGAAACATACCGATGTGTTCAACTGGA 763
Db      121 CCACCTGGGTCTCAGCGAGCCCACTGTTTCAGGAAACATACCGATGTGTTCAACTGGA 180

```

```

Qy      764 AGTGGGGGAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGGTGTGGCCAAAGATCA 823
Db      181 AGTGGGGGAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGGTGTGGCCAAAGATCA 240
Qy      824 TCATCATTTGAATCAACCCCATGTACCCAAAGACATGACATCGCCTCATGAAGTGC 883
Db      241 TCATCATTTGAATCAACCCCATGTACCCAAAGACATGACATCGCCTCATGAAGTGC 300
Qy      884 AGTTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGTGCCCTTTTGTGATGAGG 943
Db      301 AGTTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGTGCCCTTTTGTGATGAGG 360
Qy      944 AGCTCACTCCAGCACCCCACTCTGGATCACTTGGATGGGGCTTTAGGAAGCAATGGAG 1003
Db      361 AGCTCACTCCAGCACCCCACTCTGGATCACTTGGATGGGGCTTTAGGAAGCAATGGAG 420
Qy      1004 GGAAGATGTCTGACATACTCTGTCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCA 1063
Db      421 GGAAGATGTCTGACATACTCTGTCAGGCGTCAGTCCAGGTCATTGACAGCACACGGTGCA 480
Qy      1064 ATCCAGACGATGGTACCAGGGGGAAGTCAACGAGAAGATGATGTGTGAGGATCCCG 1123
Db      481 ATCCAGACGATGGTACCAGGGGGAAGTCAACGAGAAGATGATGTGTGAGGATCCCG 540
Qy      1124 AAGGGGGTGTGGACACCTGCCAGGTCGACAGTGGTGGGCCCTGATGTACCAATCTGACC 1183
Db      541 AAGGGGGTGTGGACACCTGCCAGGTCGACAGTGGTGGGCCCTGATGTACCAATCTGACC 600
Qy      1184 AGTGGCATGTGGTGGGATCGTTAGTGGGCTATGCTCGGGGGCCCGAGCACCCAG 1243
Db      601 AGTGGCATGTGGTGGGATCGTTAGTGGGCTATGCTCGGGGG--CCGAGCACCCAGG 658
Qy      1244 GAGTATACACCAAGTCTCAGCCTA 1268
Db      659 AGTTTACCACCAAGTCTCAAGCTA 683

RESULT 8
BU522841
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BU522841 895 bp mRNA linear EST 13-SEP-2002
AGENCOURT_10158166 NCI_CGAP_Co24 Mus musculus cDNA clone
IMAGE:6529422 5', mRNA sequence.
BU522841
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation: The Cepko Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14129 row: i column: 06
High quality sequence stop: 665.
Location/Qualifiers
1..895
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6529422"
/lab_host="DH10B (T1 phage-resistant)"

FEATURES
source

```

```
/clone_lib="NCI_CGAP_Co24"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match 45.9%; Score 598.4; DB 13; Length 895;
Best Local Similarity 83.0%; Pred. No. 7.3e-136;
Matches 694; Conservative 0; Mismatches 141; Indels 1; Gaps 1;

QY 454 GTGAGATTGGCCACAGCAGTCTGGATGTGTGTAATCACAGAAACAGCCAGAG 513
Db 17 GGTGGGTCCGTCAGATCAGAACCTCCCTGTGTCTCAAGTCACAGGAACAGCCAGAA 76

QY 514 CTTGCGATCGGAACCTCAAGTGGGCCCTCTCTCAGGCTCCCTGTCTCCCTGCACTGT 573
Db 77 CTTGAGGTGAGAAATGGAAGCAGATCTCTGCTCTCAGGCTCCCTGTCTCCCTGGCTGC 136

QY 574 CTTGCTGTGGGAAGCCTGAGACCCCTCTGTGTGGTGGTGGGAGAGGCTCTGTG 633
Db 137 CTTGACTGTGGAAGAGCCTGAGACTCTCTGTGTGGTGGTGGGAGGCTCTGTG 196

QY 634 GATCTTGGCTTGGCAGTCTGAGCTACAGTACGACAAACAGCAGCTCTGTGGAGGAGC 693
Db 197 GATCTTGGCTTGGCAGTCTGAGCTACAGTACGACAAACAGCAGCTCTGTGGTGGAGC 256

QY 694 ATCTGGAACCCCACTGTGGTCTCTCAGGAGCCCACTGCTTCAAGAAACATACCGATGTG 753
Db 257 ATCTGGATCCCACTGTGGTCTCTCAGGAGCCCACTGCTTCAAGAAATATCTTATGTG 316

QY 754 TCACTGGAAGTGGGAGCAGCTCAGCAAACTGGGAGCTTCCATCCCTGGCTGTG 813
Db 317 TCACTGGAAGTGGGAGCAGCTCAGCAAACTGGGAGCTTCCATCCCTGGCTGTG 376

QY 814 GCAAGATCATCATATTGAATTAACCCCATGTACCCCAAGAAACATGATGATGCTC 873
Db 377 GCAAGATCTCATGCTCAACCCCATGTGTGTACCCCAAGAAACATGATGCTC 436

QY 874 ATGAAGTCAAGTCCCACTGCTTCTCAGGACAGTCAAGCCCATCTGTGCTCCTTC 933
Db 437 ATGAAGTCAAGTCCCACTGCTTCTCAGGACAGTCAAGCCCATCTGTGCTCCTTC 496

QY 934 TTTGATGAGAGTCACTCCAGCCACCCCACTCTGGATCATTTGAGTGGGCTTTACGAAG 993
Db 497 TTTGATGAGAGTCACTCCAGCCACCCCACTCTGGATCATTTGAGTGGGCTTTACGAAG 556

QY 994 CAGATGGAAGGAGATGTGACATATCTGTGAGGGGTGATGTCAGGTCAATGACAGC 1053
Db 557 GAAACGGGAGGAGATGTGACATATCTGTGAGGGGTGATGTCAGGTCAATGACAGC 616

QY 1054 ACACGTTGCAATGCAAGCATGCTACAGGGGGAAGTCAACGAGAAATGATGTGCA 1113
Db 617 ACACGTTGCAATGCAAGCATGCTACAGGGGGAAGTCAACGAGAAATGATGTGCA 676

QY 1114 GGCATCCCGAAGGGGGTGTGACACCTGTCAGGTTGACAGTGTGGGCCCTGTATGATAC 1173
Db 677 GGCATCCCGAAGGGGGTGTGACACCTGTCAGGTTGACAGTGTGGGCCCTGTATGATAC 736

QY 1174 CAATCTGACAGTGGCATGTGTGGGCTGCTAGTGGGCTATGCTGGGGGCGCG 1233
Db 737 CAATCTGACAGTGGCATGTGTGGGCTGCTAGTGGGCTATGCTGGGGGCGCG 796

QY 1234 AGCAC-CCGAGGATATACCAAGTCTTACGCTTATCTCAACTGGATCTCAATG 1288
Db 797 AGTACTCCCTGGAGTGTATACCAAGTCTTACGCTTATCTCAACTGGATCTCAATG 852
```

RESULT 9

BM795149

LOCUS

DEFINITION X-EST0076707 S21SNUS20 Homo sapiens cDNA clone S21SNUS20-27-H07 5', mRNA sequence.

```
BM795149
BM795149.1 - GI:19143381
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 617)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 27 row: H column: 07
High quality sequence stop: 617.
Location/Qualifiers
1..617
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S21SNUS20-27-H07"
/sex="F"
/tissue_type="Stomach"
/cell_line="SNU-520"
/lab_host="Top10F"
/clone_lib="S21SNUS20"
/notes="Organ: Stomach; Vector: pTZ19RPI; Site: 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
```

ORIGIN

```
Query Match 45.0%; Score 587.8; DB 12; Length 617;
Best Local Similarity 98.1%; Pred. No. 2.3e-133;
Matches 605; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 225 TGACGAGAGCTGGATGTCTCCCTTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCC 284
Db 1 TGACGAGAGCTGGATGTCTCCCTTGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCC 60

QY 285 CGAAGGCTTCAGTGGCAGTCCGCTCTCCAGGAGCCGATCCACTCCAGTGTCTGGA 344
Db 61 CGAAGGCTTCAGTGGCAGTCCGCTCTCCAGGAGCCGATCCACTCCAGTGTCTGGA 120

QY 345 CTCGGCCACAGGGAAGTGTCTCTCCCTGTTCGACAACTTCACAGAGCTCTCGCTGA 404
Db 121 CTCGGCCACAGGGAAGTGTCTCTCCCTGTTCGACAACTTCACAGAGCTCTCGCTGA 180

QY 405 GACAGCTGTAGCAGATGGCTACAGCAGCAACCCACTTTTCAGAGCTGTGGAGTTGG 464
Db 181 GACAGCTGTAGCAGATGGCTACAGCAGCAACCCACTTTTCAGAGCTGTGGAGTTGG 240

QY 465 CCCAGCAGGAGTCTGGATGTTGTTGAAATCACAGAAAACAGCCAGGAGCTTCGATGCG 524
```

```

Db      241  CCACAGCAGGATCTGGATGTGTTGAAATCACAGAAACAGCCAGGAGCTTCGCATCG 300
QY      525  GAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGCTCTCCCTGCACTGCTCTGCTCTGG 584
Db      301  GAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGCTCTCCCTGCACTGCTCTGCTCTGG 360
QY      585  GAAGAGCCTGAAGACCCCGCTGTGTGTGGGTGGG---AGGAGGCTCTCTGTGGATTCTTG 641
Db      361  GAAGAGCCTGAAGACCCCGCTGTGTGTGGGTGGGAGGAGCCCTCTGTGGATTCTTG 420
QY      642  GCCTTGGGAGGTCAAGATCAGTACAGCAAAACAGACAGTCTGTGTGGAGGAGCATCTTGA 701
Db      421  GCCTTGGGAGGTCAAGATCAGTACAGCAAAACAGACAGTCTGTGTGGAGGAGCATCTTGA 480
QY      702  CCCCACCTGGGTCCTCAGGAGCCCACTGTCTTCAAGGAAACATACCGATGTGTTCAACTG 761
Db      481  CCCCACCTGGTCTTCAAGGAGCCCACTGTCTTCAAGGAAACATACCGATGTGTTCAACTG 540
QY      762  GAAGTGGGCGGAGGCTCAGACAAACTGGGAGCTTCCCATCCCTGGTGTGGCCAAAGAT 821
Db      541  GAAGTGGGCGGAGGCTCAAAACAAACTGGGAGCTTCCCATCCCTGGTGTGGCCAAAGAT 600
QY      822  CATCATCATTTGAATTCA 838
Db      601  CATCATCATTTGAATTCA 617

```

RESULT 10

```

LOCUS      BE616663
DEFINITION 60127896F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3610990 5',
            mRNA sequence.

```

ACCESSION

VERSION BE616663

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1. (bases 1 to 920)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabps-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLC267 row: o column: 23

High quality sequence start: 5

High quality sequence stop: 819.

Location/Qualifiers

1. 920

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3610990"

/tissue type="adenocarcinoma"

/lab host="DH10B (phage-resistant)"

/clone.lib="NIH_MGC_39"

/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI;

Site: 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed

by Ling Hong in the laboratory of Gerald M. Rubin

(University of California, Berkeley) using ZAP-cDNA

synthesis kit (Stratagene) and Superscript II RT (Life

Technologies)."

FEATURES

source

ORIGIN

Query Match 44.8%; Score 584.6; DB 10; Length 920;
 Best Local Similarity 90.4%; Pred. No. 1.8e-132;
 Matches 784; Conservative 0; Mismatches 59; Indels 24; Gaps 14;

```

QY      114  ACTCAGGCTGGCGA-GTATCATATTGTGTGTGTCTTCATCAAGGTGATTCTGGATAAAT 172
Db      14   AATGAGCCTGGGAGGTATCATATTGTGTGTCTTCATCAAGGTGATTCTGGATAAAT 73
QY      173  ACTACTTCTCTCTGGGCGAGCCTCTCCACTTCATPCCGAGGAAGCAGCTGTGTGACGGAG 232
Db      74   ACTACTTCTCTCTGGGCGAGCCTCTCCACTTCATPCCGAGGAAGCAGCTGTGTGACGGAG 133
QY      233  AGCTGGAGCTGTCCCTTGGGCGAGGAGCAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGGC 292
Db      134  AGCTGGAGCTGTCCCTTGGGCGAGGAGCAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGGC 193
QY      293  CTGAGTGGCAGTCCGCTCTCCAA--GGACCGATCCCACTGACAGTGTGTGAGCTCGGCC 351
Db      194  CTGAGTGGCAGTCCGCTCTCCAAAGGGACCGATCCCACTGACAGTGTGTGAGCTCGGCC 253
QY      352  ACAGGGAACCTGTCTCTGCTGCTTTTCGAACTTTCACAGAACTTCGCTGTGAGACAGCC 411
Db      254  ACAGGGAACCTGTCTCTGCTGCTTTTCGAACTTTCACAGAACTTCGCTGTGAGACAGCC 313
QY      412  TGTAGGCAGATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGAC 471
Db      314  TGTAGGCAGATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGAC 373
QY      472  CAGGATCTGATGTGTGTGAAATCACAGAAAACAGCCAGGAGCTTCGATCGGAACTCA 531
Db      374  CAGGATCTGATGTGTGTGAAATCACAGAAAACAGCCAGGAGCTTCGATCGGAACTCA 433
QY      532  AGTGGGCGCTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGG--AAGAG 590
Db      434  AGTGGGCGCTGTCTCTCAGGCTCCCTGGTCTCCCTGCACCTGTCTGCTGTGGGAAAGAG 493
QY      591  CTTGAAGACCCCGCTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTTGGCC--TTGGC 649
Db      494  CTTGAAGACCCCGCTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTTGGCCTTTGGC 553
QY      650  AGGTGAGCATCCAGTA-CGACAAACAGCAGTCTGTGTGGAGGAGCATCTCGACCCCCAC 708
Db      554  AGGTGAGCATCCAGTACCGAACAAACAGCAGTCTGTGTGGAGGAGCATCTCGGA--CCCC 612
QY      709  TGGGTCTCTACGGCAGCCCACTGCTTTCAGAAACATACCGA---TGTGTTCAACTGGAAG 765
Db      613  TGGGTCTCTACGGAAG-CCACTGTTTTCGGAACATACCGAATGTTGTTCCACTTGAAG 671
QY      766  GTG-CGGGCGAGGCTCA-----GACAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCAAG 819
Db      672  GTGCGGGCGAGGCTCAAGAACAAACCTGGGCAAGGTTCCAATACCTGTGTGTGGTGGCCA 731
QY      820  ATCATCAT-----CATTTGAATTCAACCCCATGTACCCCAAGACAA--TGACATCGCCCTC 873
Db      732  AAGATCCTTCATCATTTGAATTCAACCCCATGTACCCCAAGACAAATTGACATCGCCCTC 791
QY      874  ATGAAGCTGAGTTCCTCCACTCACTTTCTCAGGCACAGTCAAGCCCATCTGTGTGCCCTTC 933
Db      792  ATGAAGCTGAAGTT-CCACTCACTTTCTCAGGCACCAAGAGG-CCATCTGTTTGCCTCTT 849
QY      934  TTTGATGAGGAGTCACTCCAGCCACC 960
Db      850  GATGAGGAGTCCATCCAGCACCCCATC 876

```

RESULT 11

LOCUS BE622914

DEFINITION

602727917F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4867108 5',

mRNA sequence.

ACCESSION

BE622914

BE622914 804 bp mRNA linear EST 22-MAY-2001
 602727917F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4867108 5',
 mRNA sequence.
 BE622914

VERSION	BG822914.1	GI:14170501	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	NIH-MGC http://mgs.nci.nih.gov/		
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)		
TITLE	Unpublished (1999)		
JOURNAL	Contact: Robert Strausberg, Ph.D.		
COMMENT	Email: cgabbs-remail.nih.gov		
	Tissue procurement: ATCC		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLCW1735 row: b column: 05		
	High quality sequence stop: 793.		
FEATURES	Location/Qualifiers		
source	1..804		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4867108"		
	/tissue_type="adenocarcinoma cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_15"		
	/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:		
	ECORI; cDNA made by oligo-dT priming. Directionally		
	cloned into ECORI/XhoI sites using the following 5'		
	adaptor: GGACAGCAG(G). Size-selected >500bp for average		
	insert size 1.8kb. Library constructed by Ling Hong in		
	the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies)"		
ORIGIN			
Query Match	42.2%;	Score 551.2;	DB 12; Length 804;
Best Local Similarity	99.0%;	Pred. No. 2.7e-124;	
Matches	586;	Conservative 0;	Mismatches 3; Indels 3; Gaps 3;
QY	1	ATGGATCTCTGACAGTATCAACTCTGAAACAGCTCTGATGTCAAAACCCCTGGCGAAACCC	60
DB	215	ATGGATCTCTGACAGTATCAACTCTGAAACAGCTCTGATGTCAAAACCCCTGGCGAAACCC	274
QY	61	CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC	120
DB	275	CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC	334
QY	121	CTGGCGAGTATCATCATCTGTTGTCTTCATCAAGGTGATCTCGATAAATACTACTTC	180
DB	335	CTGGCGAGTATCATCATCTGTTGTCTTCATCAAGGTGATCTCGATAAATACTACTTC	394
QY	181	CTCTGGGGAGAGCTTCACCTTCATCCGAGGAGCAGCTGTGACGAGAGCTGGAC	240
DB	395	CTCTGGGGAGAGCTTCACCTTCATCCGAGGAGCAGCTGTGACGAGAGCTGGAC	454
QY	241	TGTCCCTTGGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTG	300
DB	455	TGTCCCTTGGGGAGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTG	514
QY	301	CGAGTCCGCTCTCCAGAGCCGATCCACATGTCAGTCTCTGGACTCGGCCACAGGAAAC	360
DB	515	GCAGTCCGCTCTCCAGAGCCGATCCACATGTCAGTCTCTGGACTCGGCCACAGGAAAC	574
QY	361	TGTTCTCTCGCTGTGTTTCAGCAACTTTCACAG-AAGCTCTCGCTGAGACAGCCTGTAGGCA	419
DB	575	TGTTCTCTCGCTGTGTTTCAGCAACTTTCACAGAAAGCTCTCGCTGAGACAGCCTGTAGGCA	634
QY	420	GATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGCAGGATCT	479
Db	635	GATGGGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCAGCAGGATCT	694
QY	480	GGATGTTGTGAATCACAGAAACAGCAGGAGCTTCGCATGCGGAACCTCAAGTGGGC	539
DB	695	GGATGTTGTGAATCACAGAAACAGCAGGAGCTTCGCATGCGGAACCTCAAGTGGGC	753
QY	540	CTCTCTCTCAGGCTCCCTGCTCTCCCTGCACACTGTCTTGCTGTGGGAAGC	591
DB	754	CTCTCTCTCAGGCT-CTGCTCTCCCTGAACCTGTCTTGCTGTGGGAAGC	804
RESULT 12			
LOCUS	BM686292	569 bp	mRNA linear EST 28-FEB-2002
DEFINITION	UI-E-CRO-adj-g-09-0-UI.r1 UI-E-CRO Homo sapiens cDNA clone		
ACCESSION	BM686292		
VERSION	BM686292.1	GI:18999550	
KEYWORDS	EST.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens (human)		
REFERENCE	1 (bases 1 to 569)		
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.		
TITLE	Normalization and subtraction: two approaches to facilitate gene		
JOURNAL	discovery		
MEDLINE	Genome Res. 6 (9), 791-806 (1996)		
COMMENT	97044477		
	8889548		
	Contact: Soares, MB		
	Coordinated Laboratory for Computational Genomics		
	University of Iowa		
	375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA		
	Tel: 319 335 8250		
	Fax: 319 335 9565		
	Email: bento-soares@uiowa.edu		
	Tissue Procurement: Dr. Gregg Hageman		
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: Researchers may obtain clones from Research		
	Genetics (www.resgen.com).		
	Seq primer: M13 Reverse.		
FEATURES	Location/Qualifiers		
source	1..569		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="UI-E-CRO-adj-g-09-0-UI"		
	/tissue_type="eye anterior segment"		
	/dev_stage="adult"		
	/lab_host="UI-E-CRO"		
	/clone_lib="UI-E-CRO"		
	/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a		
	modified polylinker; Site_1: EcoR I; Site_2: Not I;		
	UI-E-CRO is a cDNA library containing the following		
	tissue(s): eye anterior segment. The library was		
	constructed according to Bonaldo, Lennon and Soares,		
	Genome Research, 6:791-806, 1996. First strand cDNA,		
	synthesis was primed with an oligo-dT primer containing a		
	Not I site. Double stranded cDNA was ligated to an EcoR I		
	adaptor, digested with Not I, and cloned directionally		
	into pT73-Pac vector. The oligonucleotide used to prime		
	the synthesis of first-strand cDNA contains a library tag		
	sequence that is located between the Not I site and the		
	(d)18 tail. The sequence tag for this library is		
	AATGCCGCAAT. This library was created for the program, Gene		
	Discovery in the Visual System, supported by National Eye		
	Institute (NEI)."		
ORIGIN			


```

RESULT 14
BU523218
LOCUS
DEFINITION
AGENCOURT 10154487 NCI CGAP Co24 Mus musculus CDNA clone
IMAGE:6523864 5', mRNA sequence.
ACCESSION
BU523218
VERSION
BU523218.1 GI:22833656
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 956)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL4130 row: k column: 16
High quality sequence stop: 706.
Location/Qualifiers
1..956
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6523864"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Co24"
/notes="Organ: Colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
ORIGIN
Query Match 40.5%; Score 528.2; DB 13; Length 956;
Best Local Similarity 81.2%; Pred. No. 1.4e-118;
Matches 662; Conservative 0; Mismatches 148; Indels 5; Gaps 4;
QY 278 GCTTCCCGAAGGCTGCAGTGGGAGTCCGCTCTCCAGGACCGATCCACATGACAGG 337
DB 38 GCGTCGTAAGAACCGGAGTGGGAGTCCGCTCTCCAGGACAGATCCACCTGACAG 97
QY 338 TGTGACTCGGCGACAGGAACTGTCTCTGCTCTGTTTCGACAACTTCACAGAAGTCC 397
DB 98 TCGTGATCAGCCACAGGAGCTGGGCTCAGTCTGTTTCGACAACTTCACAGAAGCAC 157
QY 398 TCGTGAGACAGCTGTAGGAGATGGGTACAGGACAAACCCACTTCAGAGCTGTGG 457
DB 158 TGCCCAAGACAGCTGCACAGATGGGTATGACAGGACCGCTTCACAGAGTGG 217
QY 458 AGATTCGCCACAGGATCGGATGTTGTTGAATACAGAAACACGACGAGGAGCTTC 517
DB 218 AGATCCGTCAGATCAGACCTCCCTGTTGCTCAAGTCACAGGAACAGGAGGAACTTC 277
QY 518 GCATCGGAACCTCAAGTGGGCTGTCTCTCAGGCTCCCTGTCTCCCTGCACTCTCTTG 577
DB 278 AGGTGAGAAAGAGGAGCAGATCTGCTCTCAGGCTCCCTGTCTCTGCTGCTCCCTTG 337
QY 578 CTGTGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 637
DB 338 ACTGTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 397
QY 638 CTTGGCCTTGGCAGGTCAGATCCAGTACGACAAACAGACACGTCGTGTGTGGAGGAGCATCC 697

```

```

Db 398 CTTGGCCGTGGCAGGTTCAGCATCCAGTACAAAGCAGCATGCTGTGTGGAGCATCC 457
QY 698 TGGACCCCTCAGTGGTCTTCACGGCAGCCACTGTCTCAGGAAACATACCGATGTGTCA 757
Db 458 TGGATCCCTCAGTGGATCTTCACAGAGCCCACTGCTTCAGGAAATATCTTGAATGTCAA 517
QY 758 ACTGGAAGGTGGGGCAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGCTGTGGGCA 817
Db 518 GCTGGAAGGTTCAGGCGAGGCTCAAACTACTGGTAACTCTCCATCTTGCCTGTGGCCA 577
QY 818 AGATCATCATCATGATTCATCAACCCCATGTATCCCAAGACAAATGACATCGCCCTCATGA 877
Db 578 AGATCTTCATCTGCTGAACCCATCTCTGTATCCCAAGAGAGGACATTCGCCCTTGTGA 637
QY 878 AGCTGAGTTCCTCAGTCACTTCTCTCAGGCACAGTTCAGGCCCATCTGTGTGCCCTTCTTG 937
Db 638 AGCTGAGATGCCACTCACAATCTCAGGCTCAGTCAGGCCCATCTGCTGCCCTTCTCTG 697
QY 938 ATGAGAGTCACTCCAGCCACCCCACTCTGATCATTTGGATGGG--CTTTACGAAGCA 995
Db 698 ATGAGGTGCTTGTCCAGCCACACACCATCTGGGTCTATTTGGATGGGCTTTACGAAGAA 757
QY 996 GAATGAGGAGATCTCTGACATATCTGTC-AGGCGTCAGT-CCAGGTCAATTCAGCAGC 1053
Db 758 AACCGAGGAGATGTCCTGACATGCTACTGCAAGGCATCAGTCCAGGTCATCCAGGTC 817
QY 1054 ACACGGT-GCAATGCAGACGATCGTACCAGGGG 1087
Db 818 ACACGGTGGCAATGCAGAGGATGCTTACCAAGGG 852

RESULT 15
BE615750
LOCUS
DEFINITION
601279885F1 NIH_MGC_39 Homo sapiens cdna clone IMAGE:3621980 5',
mRNA sequence.
ACCESSION
BE615750
VERSION
BE615750.1 GI:9897349
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 787)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM296 row: i column: 21
High quality sequence stop: 690.
Location/Qualifiers
1..787
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3621980"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_39"
/notes="Organ: pancreas; Vector: pOT7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin

```

(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."

ORIGIN

```

Query Match      39.4%; Score 513.6; DB 10; Length 787;
Best Local Similarity 96.2%; Pred. No. 4.7e-115;
Matches 558; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCCCTGATGTCAAAACCCCTGGCAAAACCCCG 62
Db 210 GGATCCTGACAGTATCAACCTCTGAACAGCCCTGATGTCAAAACCCCTGGCAAAACCCCG 269
QY 63 TATCCCCATGGAGACCTTCAGAAA-GGTGGGGATCCCCATCATCATAGCACTACTGAGCC 121
Db 270 TATCCCCATGGAGACCTTCAGAAAATGGTGGGATCCCCATCATCATAGCACTACTGAGCC 329
QY 122 TGGCAGATATCATATTGTTGGTGTCTCATCAAGTGTATTCTGGATAAATACTACTTCC 181
Db 330 TGGCAGATATCATATTGTTGGTGTCTCATCAAGTGTATTCTGGATAAATACTACTTCC 389
QY 182 TCTGGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTACGGAGCTGGACT 241
Db 390 TCTGGGGCAGCCTCTCCACTTCATCCCGAGGAAGCAGCTGTGTACGGAGCTGGACT 449
QY 242 GTCCCTTGGGGAGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTGCAGTGG 301
Db 450 GTCCCTTGGGGAGACGAGGAGCACTGTGTCAAGAGCTTCCCCGAAGGGCTGCAGTGG 509
QY 302 CAGTCCGCTCTCCAAGACCGATCCACACTGCGAGGTCTGGACTCGGCCACAGGAACT 361
Db 510 CAGTCCGCTCTCCAAGACCGATCCACACTGCGAGGTCTGGACTCGGCCACAGGAACT 569
QY 362 GGTCTCTCGCTGTTCGACAACTTCACAGAGCTCTCGTGAGACAGCTGTAGGCAGA 421
Db 570 GGTCTCTCGCTGTTCGACAACTTCACAGAGCTCTCGTGAGACAGCTGTAGGCAGA 629
QY 422 TGGGCTACAGCAGCAAAACCACTTTACAGAGCTGTGGAGATTGGCCACAGCAGGATCTGG 481
Db 630 TGGGCTACAGCAGCAAAACCACTTTACAGAGCTGTGGAGATTGGCCACAGCAGGATCTGG 689
QY 482 ATGTTGTTGAATCAGAAAACAGCCAGGAGCTTCGCATGGGNACTCAAGTGGGCCCT 541
Db 690 ATGTCGACGACATCAGAAAACAGCCAGGAGCTTCGCCTTGGGACTCAAGTGGG-CCT 748
QY 542 GTCTCTCAGGCTCCCTGTCTCCCTGCACCTGTCTTGCCTG 581
Db 749 GTCTCTCAGGCTCCCTGGGTCT-CCTGCACCTGTCTTGA CTG 787

```

Search completed: June 8, 2004, 02:20:27
Job time : 3787 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 7, 2004, 23:24:49 ; Search time 113 Seconds
(without alignments)
6408.946 Million cell updates/sec

Title: US-10-030-688-1

Perfect score: 1305

Sequence: 1 atgagctctacagctgatac.....atgtctggaaggctgagctg 1305

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgm2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgm2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgm2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgm2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgm2_6/ptodata/2/ina/PTCUS COMB.seq.*

6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1303.4	99.9	2038	3	US-09-008-271A-18
2	1303	99.8	2081	4	US-09-851-588-7
3	1292	99.0	2079	4	US-09-656-002-1
4	1292	99.0	2079	4	US-09-851-588-5
5	1255	11.9	2413	3	US-09-518-046-1
6	150	11.5	1479	3	US-09-342-749-1
7	150	11.5	1479	4	US-09-691-840-1
8	147	11.3	2479	3	US-09-342-749-29
9	147	11.3	2479	4	US-09-691-840-29
10	147	11.3	2479	4	US-09-685-166A-894
11	142.8	10.9	1077	3	US-08-807-151-2
12	142.8	10.9	1077	4	US-09-478-957-2
13	133.6	10.2	2416	3	US-09-261-416-1
14	132.8	9.4	2544	3	US-09-518-046-3
15	119.4	9.1	942	4	US-09-636-382A-3
16	119.2	9.1	1154	4	US-09-636-382A-1
17	112	8.6	1110	4	US-09-386-653A-1
18	112	8.6	1212	4	US-09-620-312D-431
19	111.6	8.6	696	1	US-08-508-448C-24
20	111.6	8.6	1460	4	US-09-370-838-80
21	111.6	8.6	1517	1	US-08-508-448C-15
22	111.6	8.6	2790	4	US-09-370-838-79
23	111	8.5	980	4	US-09-023-942A-30
24	110.2	8.4	901	1	US-08-508-448C-9
25	108.4	8.3	1462	4	US-09-370-838-55
26	106.4	8.2	1378	4	US-09-907-794A-262
27	106.4	8.2	1378	4	US-09-905-125A-262

28 106.4 8.2 1378 4 US-09-902-775A-262
29 106.4 8.2 1430 4 US-09-386-629-1
30 99.8 7.6 1440 1 US-07-882-202A-3
31 99.8 7.6 1440 1 US-08-021-615A-3
32 99.8 7.6 1440 1 US-08-321-777-3
33 99.8 7.6 1440 3 US-09-009-217-13
34 99.8 7.6 1440 3 US-09-009-656-13
35 99.8 7.6 1440 5 PCT-US93-04493-3
36 99.8 7.6 2462 2 US-08-479-733A-25
37 99.8 7.6 2462 3 US-08-487-427-25
38 99.8 7.6 2462 3 US-08-479-727A-25
39 99.8 7.6 2462 3 US-08-482-369A-25
40 99.8 7.6 2462 5 PCT-US95-07439-25
41 98.2 7.5 2422 1 US-08-475-845-1
42 98.2 7.5 2422 1 US-08-327-690-1
43 98.2 7.5 2422 2 US-08-660-289-1
44 98.2 7.5 2422 2 US-08-537-807-1
45 98.2 7.5 2422 2 US-08-871-003-1

ALIGNMENTS

RESULT 1

; Sequence 18, Application US/09008271A
; Patent No. 6203979

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; Yue, Henry

; Guegler, Karl J.

; Corley, Neil C.

; Tang, Tom Y.

; Shah, Purvi

; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/008,271A

; FILING DATE: 16-Jan-1998

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: <Unknown>

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Mohan-Peterson, Sheela

; REGISTRATION NUMBER: 41,201

; REFERENCE/DOCKET NUMBER: PF-0458 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNOT13

; CLONE: 1337018

; SEQUENCE DESCRIPTION: SEQ ID NO: 18:

; US-09-008-271A-18

```
Query Match      99.9%; Score 1303.4; DB 3; Length 2081;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGATCCTGACAGTATCACTCTGAAACAGCCTCGATGTCAAAACCCCTGCGCAACCC 60
DB 200 ATGGATCCTGACAGTATCACTCTGAAACAGCCTCGATGTCAAAACCCCTGCGCAACCC 259
QY 61 CGTATCCCATGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 120
DB 260 CGTATCCCATGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGC 319
QY 121 CTGGGAGTATCATCATTTGGTGTCTCTCATCAAGTGTATCTGGATAAATACTACTTC 180
DB 320 CTGGGAGTATCATCATTTGGTGTCTCTCATCAAGTGTATCTGGATAAATACTACTTC 379
QY 181 CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 240
DB 380 CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGAC 439
QY 241 TGTCTCTTGGGGAGGACGAGAGACATGTGTCAAGAGCTTCCCGAAGGCCCTGCGAGTG 300
DB 440 TGTCTCTTGGGGAGGACGAGAGACATGTGTCAAGAGCTTCCCGAAGGCCCTGCGAGTG 499
QY 301 CGAGTCCGCTCTCCAAGGACCGATCCACATCTGCAAGTGTCTGGACTGGCCACAGGGAAC 360
DB 500 CGAGTCCGCTCTCCAAGGACCGATCCACATCTGCAAGTGTCTGGACTGGCCACAGGGAAC 559
QY 361 TGGTCTCTGCTGTTTGGCAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGGCAG 420
DB 560 TGGTCTCTGCTGTTTGGCAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGGCAG 619
QY 421 ATGGGCTACAGCAGAAACCCACTTTCAGAGCTGTGGAGATGGCCAGACAGAGATCTG 480
DB 620 ATGGGCTACAGCAGAAACCCACTTTCAGAGCTGTGGAGATGGCCAGACAGAGATCTG 679
QY 481 GATGTGTTGTAATCACAGAAACAGCAGGAGCTTCGATCGCGAACTCAAGTGGGCCC 540
DB 680 GATGTGTTGTAATCACAGAAACAGCAGGAGCTTCGATCGCGAACTCAAGTGGGCCC 739
QY 541 TGTCTCTCAGGCTCCTGGTCTCCCTGCACTGTCTTGTCTGTGGAGAGAGCTGAAAGCC 600
DB 740 TGTCTCTCAGGCTCCTGGTCTCCCTGCACTGTCTTGTCTGTGGAGAGAGCTGAAAGCC 799
QY 601 CCCGCTGTGGTGGGGAGAGGAGCCTCTGTGGATCTTGGCTTGGCAGGTGAGCATC 660
DB 800 CCCGCTGTGGTGGGGAGAGGAGCCTCTGTGGATCTTGGCTTGGCAGGTGAGCATC 859
QY 661 CAGTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTGGAACCCCACTGGGTCTCAAG 720
DB 860 CAGTACGCAAAACAGCAGCTCTGTGGAGGAGCATCTGGAACCCCACTGGGTCTCAAG 919
QY 721 CGAGCCCACTGCTTCAGGAACATACCGATGTGTCACTGAGAGGTGGGGCAGGCTCA 780
DB 920 CGAGCCCACTGCTTCAGGAACATACCGATGTGTCACTGAGAGGTGGGGCAGGCTCA 979
QY 781 GACAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTC AAC 840
DB 980 GACAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATTTGAATTC AAC 1039
QY 841 CCATGTATCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTC 900
DB 1040 CCATGTATCCCAAGACATGACATCGCCCTCATGAAGCTGCAGTTCCCACTCACTTTC 1099
QY 901 TCAGGCACAGTCAGGCCCATCTGTCTGCTGCTTCTTGTATGAGGAGCTCACTCCAGCCACC 960
DB 1100 TCAGGCACAGTCAGGCCCATCTGTCTGCTGCTTCTTGTATGAGGAGCTCACTCCAGCCACC 1159
QY 961 CCACCTCTGGATCATTTGGATGGGGCTTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATA 1020
DB 1160 CCACCTCTGGATCATTTGGATGGGGCTTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATA 1219

RESULT 2
US-09-851-588-7
; Sequence 7, Application US/09851588
; Patent No. 6682890
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2081
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (215)..(1528)
; OTHER INFORMATION:
US-09-851-588-7

Query Match      99.8%; Score 1303; DB 4; Length 2081;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCG 62
DB 223 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAACCCCG 282
QY 63 TATCCCATGGAGACCTTCAGAAAGTGGGGATTCCTCATCATAGCACTACTAGAGCT 122
DB 283 TATCCCATGGAGACCTTCAGAAAGTGGGGATTCCTCATCATAGCACTACTAGAGCT 342
QY 123 GGGAGTATCATCATTTGGTGTCTCTCATCAAGTGTATCTGGATAAATACTACTTCT 182
DB 343 GGGAGTATCATCATTTGGTGTCTCTCATCAAGTGTATCTGGATAAATACTACTTCT 402
QY 183 CTGCGGGAGCCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGAGTG 242
DB 403 CTGCGGGAGCCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGAGTG 462
QY 243 TCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGAGGCTGAGTG 302
DB 463 TCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGAGGCTGAGTG 522
```

QY 303 AGTCGGCTCTCCAGGACCGATCCACACTGCTGAGTCTGACTCGGCGCACAGGAACTG 362
DB 523 AGTCGGCTCTCCAGGACCGATCCACACTGCTGAGTCTGACTCGGCGCACAGGAACTG 582
QY 363 GTTCTCTGCTCTGTTTGGCAACTTTCAGAAAGCTCTGCTGAGACAGCTGTAGGAGAT 422
DB 583 GTTCTCTGCTCTGTTTGGCAACTTTCAGAAAGCTCTGCTGAGACAGCTGTAGGAGAT 642
QY 423 GGGCTACAGCAGCAAAACCCATTTCAGAGCTGTGGAGATTGGCCCGACAGGATCTGGA 482
DB 643 GGGCTACAGCAGCAAAACCCATTTCAGAGCTGTGGAGATTGGCCCGACAGGATCTGGA 702
QY 483 TGTGTGTTGAAATCACAGAAAAACAGCAGGAGCTTCGATGCGGAATCAAGTGGGCCCTG 542
DB 703 TGTGTGTTGAAATCACAGAAAAACAGCAGGAGCTTCGATGCGGAATCAAGTGGGCCCTG 762
QY 543 TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTGCTGTGGAGAGAGCTGTGAAGACCC 602
DB 763 TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTGCTGTGGAGAGAGCTGTGAAGACCC 822
QY 603 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTCAGCATCCA 662
DB 823 CCGTGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCTTGGCAGGTCAGCATCCA 882
QY 663 GTACGACAAACAGCAAGCTCTGTGAGGAGAGCATCTGCAACCCCACTGGTCTCACGGC 722
DB 883 GTACGACAAACAGCAAGCTCTGTGAGGAGAGCATCTGCAACCCCACTGGTCTCACGGC 942
QY 723 AGCCCACTGCTTCAGGAACATACCGATGCTTCACTGGAGGCTGGGCGAGCTCAGA 782
DB 943 AGCCCACTGCTTCAGGAACATACCGATGCTTCACTGGAGGCTGGGCGAGCTCAGA 1002
QY 783 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACACC 842
DB 1003 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACACC 1062
QY 843 CATGTACCCCAAGACATGATCGCCCTCATGAAGCTGAGTTCGCACTCATCTTCTC 902
DB 1063 CATGTACCCCAAGACATGATCGCCCTCATGAAGCTGAGTTCGCACTCATCTTCTC 1122
QY 903 AGGCACAGTCAGGCCCTCTGCTGCTGCTCTTGTGATGAGGAGCTCACTCCAGCCACCCC 962
DB 1123 AGGCACAGTCAGGCCCTCTGCTGCTGCTCTTGTGATGAGGAGCTCACTCCAGCCACCCC 1182
QY 963 ACTCTGATCATTTGATGGGCTTTACGAGCAGAAATGGAGGAGATGTCTGACATPACT 1022
DB 1183 ACTCTGATCATTTGATGGGCTTTACGAGCAGAAATGGAGGAGATGTCTGACATPACT 1242
QY 1023 GCTCAGGCGTCAGTCCAGGTCATTGACAGCAGCAGCTGCAATGCAGCAGATGCTACCA 1082
DB 1243 GCTCAGGCGTCAGTCCAGGTCATTGACAGCAGCAGCTGCAATGCAGCAGATGCTACCA 1302
QY 1083 GGGGGAAGTCAACGAGAAATGATGTGTGAGGATCCCGGAAGGGGTGTGACACCTG 1142
DB 1303 GGGGGAAGTCAACGAGAAATGATGTGTGAGGATCCCGGAAGGGGTGTGACACCTG 1362
QY 1143 CCAGGTCAGAGTGTGGGCCCCCTGATGTACCAATCTGACAGTGGGATGTGTGGGCA 1202
DB 1363 CCAGGTCAGAGTGTGGGCCCCCTGATGTACCAATCTGACAGTGGGATGTGTGGGCA 1422
QY 1203 GGTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCAGAGATATACACCAAGGTCTC 1262
DB 1423 GGTAGCTGGGCTATGGCTCGGGGGCCCGAGCACCCAGAGATATACACCAAGGTCTC 1482
QY 1263 AGCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1305
DB 1483 AGCTATCTCAACTGGATCTACATGTCTGGAAGGCTGAGCTG 1525

RESULT 3

US-09-656-002-1

; Sequence 1, Application US/09656002

; Patent No. 6455668

; GENERAL INFORMATION:

; APPLICANT: Mack, David

; APPLICANT: Gish, Kurt

; APPLICANT: Wilson, Keith

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND N

; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

; FILE REFERENCE: A-69108/DJB/JJD/BMS

; CURRENT APPLICATION NUMBER: US/09/656,002

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: US 09/525,993

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: US 09/493,444

; PRIOR FILING DATE: 2000-01-28

; PRIOR APPLICATION NUMBER: PCT/US 00/07044

; PRIOR FILING DATE: 2000-03-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 1

; LENGTH: 2079

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-656-002-1

Query Match 99.0%; Score 1292; DB 4; Length 2079;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 GGATCTCTGACGTGATCAACCTCTGACAGCTCGATGTCAACCCCTCGGCAAAACCCG 62
DB 216 GGATCTCTGACGTGATCAACCTCTGACAGCTCGATGTCAACCCCTCGGCAAAACCCG 275
QY 63 TATCCCATGGAGACCTTCAGAAAG-GTGGGATCCCCATCATAGACTACTAGGAC 121
DB 276 TATCCCATGGAGACCTTCAGAAAGTGTGGGATCCCCATCATAGACTACTAGGAC 335
QY 122 TGCGAGTATCATATTGTTGTTCTCTCATCAAGTGTATTCTGGTAATAATACTACTTC 181
DB 336 TGCGAGTATCATATTGTTGTTCTCTCATCAAGTGTATTCTGGTAATAATACTACTTC 395
QY 182 TCTGCGGCGAGCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACT 241
DB 396 TCTGCGGCGAGCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGACT 455
QY 242 GTCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCCGAGGCGCTGCACTGG 301
DB 456 GTCCCTTGGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCCGAGGCGCTGCACTGG 515
QY 302 CAGTCCGCTCTCCAGGAGCCGATCCACACTGAGTGTGCTGAGCTCGGCGCACAGGAACT 361
DB 516 CAGTCCGCTCTCCAGGAGCCGATCCACACTGAGTGTGCTGAGCTCGGCGCACAGGAACT 575
QY 362 GGTCTCTGCTCTGTTTCGACAACTTTCAGAAAGCTCTGCTGAGACAGCTGTAGGCAGA 421
DB 576 GGTCTCTGCTCTGTTTCGACAACTTTCAGAAAGCTCTGCTGAGACAGCTGTAGGCAGA 635
QY 422 TGGGCTTACGAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCGACAGGATCTGG 481
DB 636 TGGGCTTACGAGCAGCAAAACCCACTTTCAGAGCTGTGGAGATTGGCCCGACAGGATCTGG 695
QY 482 ATGTTCTTGAATTCAGAAAAACAGCCAGGAGCTTTCGATGCGGAATCAAGTGGGCGCT 541
DB 696 ATGTTCTTGAATTCAGAAAAACAGCCAGGAGCTTTCGATGCGGAATCAAGTGGGCGCT 755
QY 542 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTTGGGAGAGCTGTAGAGCCC 601
DB 756 GTCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTTGGGAGAGCTGTAGAGCCC 815
QY 602 CCGTGTGTGGTGGGAGGAGGCTCTGTGATTCTTGGCTTGGCAGGTCAGCATCC 661
DB 816 CCGTGTGTGGTGGGAGGAGGCTCTGTGATTCTTGGCTTGGCAGGTCAGCATCC 875
QY 662 AGTACGACAAACAGCAGCTCTCTGTGGAGGAGCATCTCGAGCCCCCAGCTCGGCTCTCACGG 721

Db 1296 AGGGGAAAGTACCGAGAGAGATGATGTGTGAGGATCCCGAGAGGGGGTGTGACACCT 1355
QY 1142 GCCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCACTGGCATCTGTGTGGCA 1201
Db 1356 GCCAGGGTGACAGTGGTGGGCCCCCTGATGTACCAATCTGACCACTGGCATCTGTGTGGCA 1415
QY 1202 TCGTTAGTGGGGCTATGGCTGCGGGGCCCGAGACACCCAGGAGTATACACCAAGTCT 1261
Db 1416 TCGTTAGTGGGGCTATGGCTGCGGGGCCCGAGACACCCAGGAGTATACACCAAGTCT 1475
QY 1262 CAGCCTATCTCACTGGATCTACCAATCTGTGGAAGGCTGAGCTG 1305
Db 1476 CAGCCTATCTCACTGGATCTACCAATCTGTGGAAGGCTGAGCTG 1519
RESULT 5
US-09-518-046-1
; Sequence 1, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518,046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 1
; LENGTH: 2413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; OTHER INFORMATION: entire cDNA sequence of TADG-12 gene
US-09-518-046-1
Query Match 11.9%; Score 155; DB 3; Length 2413;
Best Local Similarity 53.5%; Pred. No. 1.3e-33;
Matches 420; Conservative 0; Mismatches 350; Indels 15; Gaps 4;
QY 532 AGTGGGCCCTCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGTGCTGTGGGAAGC 591
Db 717 AGGAGGGATGTGCTCTGGCCACGTTGTTACCTTGGAGTGCACAGCTGTGTGCATAGA 776
QY 592 CTG---AAGACCCCGTGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTTG 648
Db 777 AGGGGCTACAGCTCAGCATCGTGGTGGAAACATGTCTTGTCTCGCAGTGGCCCTGG 836
QY 649 CAGGTCAGCATCCAGTACGACAAACAGACAGCTCTGTGGAGGAGCATCTGGACCCAC 708
Db 837 CAGGCCAGCTTCAGTTCCAGGGTACACCTGTGGGGGGCTGTGTATCATCAGCCCTG 896
QY 709 TGGTCTCTCAGGGAGGCCACTGCTT-----CAGGAAACATACCGATGTGTTCACTGG 762
Db 897 TGGATCATCTCTGTGCACATGTTTATGACTTGTACCTCCCAAGTCATGGACCATC 956
QY 763 AAGGTGGGGAGGCTCAGACAAACTGGGAGCTTCCATCCCTGGCTGTGGCCAGATC 822
Db 957 CAGGTGGGTCTAGTTTCCCTTGTGGCAATTCAGCCCTCCCACTGGTGGGAGAAGATT 1016
QY 823 ATCATCATTTGAATTCACCCCATGTATCCCAAGACATGACATCGCCCTCATGAAGCTG 882
Db 1017 GTCTACACAGCATGACAGCCAAAGAGCTGGGCAATGACATCGCCCTTATGAGCTG 1076
QY 883 CAGTCCGACTCATTCTCAGGCACAGTACAGGCCCATCTGTGTGCCCTCTTTGATGAG 942
Db 1077 GCCGGGCACTCAGTTCAATGAATGATCCAGGCTGTGTGCCCTGCCCAACTCTGAAG 1136
QY 943 GAGCTCACTCCAGCCACCCCACTCTGGATCATTTGATGGGGCTTACGAGCAGCAATGGA 1002

Db 1137 AACTTCCCGATGGAAAGTGTGCTGGAGCTCAGGATGGG---GGCCACAGAGGATGGA 1193
QY 1003 GGAAGATGTCTGACATCTGCTGAGGGCTGAGTCCAGTCCAGTTCATTCAGACAGCACGGTGC 1062
Db 1194 GGTGACGCTCCCTGCTGTAACACAGCGGCGTCCCTTTGATTTCCAAAGATCTCTG 1253
QY 1063 AATGACAGCATGCTGCTACCAAGGGGAAAGTACCGAGAAAGATGATGTGTCAGGCAATCCCG 1122
Db 1254 AACCACAGGAGCGTGTACGCTGCGGTCATCTCCCCCTCCATGCTCTGCGCGGCTACCTG 1313
QY 1123 GAAGGGGTGTGGACACCTGCCAGGGTGACAGTGGTGGGGCCCTGATGTACCAA---TCT 1179
Db 1314 ACGGGTGGCGTGAACAGCTGCCAGGGGAGACGCGGGGGCCCTGCTGTGTCAAGAGAG 1373
QY 1180 GACCACTGGCATGTGCTGGGCTCGCTAGCTGGGGCTATGCTGGGGGGCCCGACAC 1239
Db 1374 AGGCTGTGGAAGTTAGTGGGAGGACCAAGCTTTGGCATCGGCTGCGCAGAGTGAACAAG 1433
QY 1240 CAGGAGTATACCAAGGCTCTCAGGCTATCTCAACTGGATCTACATGTCTGGAAGGCT 1299
Db 1434 CTTGGGGTGTACACCGCTGTCTCCTTCTGGACTGGATCCAGCAGCAGATGGAGAGA 1493
QY 1300 GAGCT 1304
Db 1494 GACCT 1498
RESULT 6
US-09-342-749-1
; Sequence 1, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1476)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (724)
; OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (985)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1347)
; OTHER INFORMATION: Listed as C in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1466)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1471)
; OTHER INFORMATION: Listed as A in GenBank Accession No. 6166194 U75329
; FEATURE:

NAME/KEY: allele
 LOCATION: (478)
 OTHER INFORMATION: This base can be G or A with G being the more
 OTHER INFORMATION: common allele. The codon will change from Val to
 FEATURE: Met.
 NAME/KEY: allele
 LOCATION: (777)
 OTHER INFORMATION: This base can be C or T with C being the more
 OTHER INFORMATION: common allele. The codon is unaffected with both
 OTHER INFORMATION: alleles encoding Gly.
 NAME/KEY: allele
 LOCATION: (768)
 OTHER INFORMATION: This base can be C or T with C being the more
 OTHER INFORMATION: common allele. This is a silent polymorphism.
 NAME/KEY: allele
 LOCATION: (834)
 OTHER INFORMATION: This base can be C or T with C being the more
 OTHER INFORMATION: common allele. This is a silent polymorphism.
 NAME/KEY: allele
 LOCATION: (625)
 OTHER INFORMATION: This base can be T or A with T being the more
 OTHER INFORMATION: common allele. The codon will change from Phe to
 OTHER INFORMATION: Ile

US-09-342-749-1

Query Match 11.5%; Score 150; DB 3; Length 1479;

Best Local Similarity 51.7%; Pred. No. 2.6e-32;

Matches 534; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCCGCTCTCCAGGACCGATCCACATCGAGGTGCTGGACTCGGCCACAGGAACTGG 363
 DB 445 GTTCGCTCTAGGACCAAACTTCATCTCTCAGGTGTACTCATCTCAGAGAACTCTGG 504
 QY 364 TTCTCTGCTGTTTCGACAACTTCACAGAACTCTCTGCTGACAGAGCTGTAGGACATG 423
 DB 505 CACCCTGTGTGCAAGACGACTGGAACAGAACTACGGGGCGGCGCTGACGGGACATG 564
 QY 424 GCGTCAGCAGAAACCCACTTTCAGAGCTGTGGAGATGGCCAGACAGATCTGGAT 483
 DB 565 GCGTATAAGAAATAATTTTACTCTAG-CAAAGGAATAGTGGATGACAGCGGATCCACG 623
 QY 484 GTTGTGTAATCACAGAAACAGCAGAGCTTCG-----CATGCGGAATCAAG 533
 DB 624 CTTTATGAACCTGAACAAAGTGGCGGCAATGTCGATATCTATAAACTGTACCACAG 683
 QY 534 TGGGCGCTGTCTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
 DB 684 TGATGCTGCTTCTCAAAAGCAGTGGTCTTTTACGCTGTATAGCTCGGGGTCAAAT 743
 QY 594 GAAGACCCCTC-----GTGCTGTGGTGGGAGGAGGCTCTGTGGATTCTTGCC 644
 DB 744 GAATCAAGCCCGCAGACAGATGCTGGCGGCGAGAGCGGCTCCCGGGGCTGGCC 803
 QY 645 TTGGCAGGTGACATCCAGTACGACAAACAGACCTCTGTGGAGGAGCATCTTGGACCC 704
 DB 804 CTGGCAGGTGACGCTGACCTCCAGACGCTCCACGTGTGGAGGCTCCATCATCACCCC 863
 QY 705 CCACTGGGTCTCAGGAGCCCTCTGCTCAGGAACAT-----ACCATGTGTCA 757
 DB 864 CAGTGTGATCGTGTGACAGCGCCCTCTGGTGGAACAACTCTTAAACATCATGGCATG 923
 QY 758 ACTGGAAGGTGGGGCAGGCTCAGACAAACTGGGAGCTTCCATCCCTGCTGTGGCCA 817
 DB 924 GACGCAATTTGCGGGAATTTTCAGACAACTTTTCATGTTCTATGAGCGGATACCAAGT 983
 QY 818 AGATCATCATCATTAATTAACCCCATG-----ACCCGAAGACATGACATCGCCT 872
 DB 984 AGAAAAAGTGATTTCTCATCTCAAAATATAGTCCAGACCAAGAAACATGATTCGCT 1043

QY 873 CATGAAGCTGCAGTTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTT 932
 DB 1044 GATGAAGCTGCAGAGCCTCTGACCTTTCAACGACCTAGTGAACCAAGTGTGTCTGCCCAA 1103
 QY 933 CTTTGTATGAGGAGCTCACTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAA 992
 DB 1104 CCGAGGATGATGCTGCAGCCAGAACAGCTCTGTCTGGATTTCCGGGTGGGGGCCACCGA 1163
 QY 993 GCAGATGAGGAGAGATGCTGACATATCTGCTGAGGCGTCAGTCCAGGTCAATGACAG 1052
 DB 1164 GGAGAA---AGGGAAGACCTCAGAACTGCTGAACCTGCCAAGGTCTCTCTCATTTGAGAC 1220
 QY 1053 CACACGCTCAATGCAGACGATCGGTACCGAGGGGAAAGTCACCGAGAAAGATGATGTGTC 1112
 DB 1221 ACAGAGATCAACAGCAGATATGCTATGACAACTGATCACACCAAGCATGATCTGTGC 1280
 QY 1113 AGCATCCCGGAAGGGGTGTGGACACCTGCCAGGGTGACAGTGGTGGCCCTG---AT 1169
 DB 1281 CGGCTTCTCTGAGGGGAACTCGATTCTTTGCCAGGGTGACAGTGGAGGCGCTCTGTCTAC 1340
 QY 1170 GTACCAATCTGACCAAGTGGCATGTGTGGGCATCGTTAGCTGGGGCTATGGCTGGGGGG 1229
 DB 1341 TTCGAAGAACAATATCTGTGTGCTGATAGGGGATACAAAGCTGGGTCTGTGCTGTGCCAA 1400
 QY 1230 CCGAGACACCCAGGAGTATACACCAAGGTCTCAGCTTATCTCAACTGGATCTACAATGT 1289
 DB 1401 AGCTTACAGACCAAGGAGTGTACGGGAATGTGATGTTTACCGGACTGGATTATTCGACA 1460
 QY 1290 CTGGAAGGCTGA 1301
 DB 1461 AATGAGGGCAGA 1472

RESULT 7

US-09-691-840-1
 ; Sequence 1, Application US/09691840
 ; Patent No. 6444419
 ; GENERAL INFORMATION:
 ; APPLICANT: Wong, Alexander K.C.
 ; APPLICANT: Tavtigian, Sean V.
 ; APPLICANT: Teng, David H.-F.
 ; APPLICANT: Myriad Genetics, Inc.
 ; TITLE OF INVENTION: TMRSS2 is a Tumor Suppressor
 ; FILE REFERENCE: 2318-202
 ; CURRENT APPLICATION NUMBER: US/09/691,840
 ; CURRENT FILING DATE: 2000-10-18
 ; PRIOR APPLICATION NUMBER: US/09/342,749
 ; PRIOR FILING DATE: 1999-06-29
 ; PRIOR APPLICATION NUMBER: US 60/091,044
 ; PRIOR FILING DATE: 1998-06-29
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1479
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1476)
 ; NAME/KEY: conflict
 ; LOCATION: (724)
 ; OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
 ; NAME/KEY: conflict
 ; LOCATION: (985)
 ; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1347)
 ; OTHER INFORMATION: Listed as C in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1466)
 ; OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329
 ; NAME/KEY: conflict
 ; LOCATION: (1471)

OTHER INFORMATION: Listed as A in GenBank Accession No. 6444419 U75329.
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to Ile
US-09-691-840-1

Query Match 11.5%; Score 150; DB 4; Length 1479;
Best Local Similarity 51.7%; Pred.No. 2.6e-32;
Matches 534; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCCGCTCTCCAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAATCGG 363
DB 445 GTCCGCTCTCCAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAATCGG 504
QY 364 TTCCTGCTGTTTCGACAACTTCAGAACTCTCGCTGAGACAGCTGTAGGAGATG 423
DB 505 CACCTGTGTCCAGAGAGCTGGAACAGAACTACCGGGCGGGGCTTCGAGGACATG 564
QY 424 GCCTACAGCAGCAACCTTTTCAGAGCTGTGGAGTTGGCCAGACAGGATCTGGAT 483
DB 565 GCCTAAGATATTTTACTCTAG-CCAAGGATAGTGATGACAGCGATCCACAG 623
QY 484 GTTGTGAAATCAGAGAAACAGCAGGAGTTGG-CAAGGAACTCAAG 533
DB 624 CTTTATGAACTGAAACAGAGTCCGCGCAATGTGATATCTATAAAACTGTACCACAG 683
QY 534 TGGGCTCTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
DB 684 TGATGCTGTCTTCAAAGCAGTGGTTCTTTAGCTGTATAGCTTCGCGGGTCACTT 743
QY 594 GAAGACCCCTCC-CTGTGTGGTGGGAGGAGGAGGCTCTGTGGATTTCTTGCC 644
DB 744 GAACCTCAAGCCGCGCAGACGAGATGCTGGCGGCGAGAGCGCTCCCGGGGCTGGCC 803
QY 645 TTGGCAGTTCAGATCCAGTAGCAGACAGCAGTCTGTGGAGGAGCATCTCGGAGCC 704
DB 804 CTGGCAGTTCAGCTTCAGCTCCAGAACGTCCAGTGTGGAGGCTCCATCATCACCCC 863
QY 705 CCACCTGGCTCTCAGCGAGCCCTGCTTCAGGAAACAT-ACCGATGTGTCA 757
DB 864 CGAGTGGATGTGACAGCGCCCTGCTGGGAAACCTCTTAACATCCATGCCATG 923
QY 758 ACTGAAAGTGGGCGCAGGCTCAGACAACTGGGAGCTTCCCATCTCTGCTGTGGCCA 817
DB 924 GACGGCAATTCGGGGATTTTGAACAATCTTTATGTTCTATGGAGCGGATACCAAGT 983
QY 818 AGATCATCATCTGAATTCACCCCATGT-ACCCCAAGACATGACATCGCCT 872
DB 984 AGAAAAGTATTTCTCATCCAAATTATGACTCCAGACCAAGAACATGACATTTGGCT 1043
QY 873 CATGAAGTGCAGTTTCCCACTCATTCTCAGGACAGCTCAGGCCCCATCTCTGCTGCT 932
DB 1044 GATGAAGTGCAGAGGCTCTGACTTTCAAGACCTAGTGAACACAGTGTCTGTGCCAA 1103

QY 933 CTTTGTAGAGGAGCTCATTCCAGCCACCCACTCTGGATCATTTGGATGGGGCTTTACGAA 992
DB 1104 CCAGGATGATGCTCGAGCCAGAAAGCTCTGCTGATTTCCGGTGGGGGCCACCGA 1163
QY 993 GCAGATGAGGAGGAGATGCTGACATATCTGTGAGGCGTCTGAGTCCAGTCAATGACAG 1052
DB 1164 GGAGAA-AGGGAAGACCTCAGAGTGTGAACGCTGCAAGTCTTCTCATTTAGAC 1220
QY 1053 CACACGCTGCAATGACAGACATGCTGACAGGGGAAAGTCCACGAGAAAGATGATGTGC 1112
DB 1221 ACAGAGATGCAACAGCAGATATGTCTATGACAACTGTATGACACACAGCCATGATCTGTGC 1280
QY 1113 AGCATCCCGGAGGGGTGTGGACACCTCCAGGCTGACAGTGTGGGCGCTG-AT 1169
DB 1281 CGGCTTCTCGAGGGGAAGCTGATTTCTTGCCAGGCTGACAGTGGAGGCTCTGTGCTAC 1340
QY 1170 GTACCAATCTGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1229
DB 1341 TTCAAGAACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1400
QY 1230 CCGGACACCCAGGAGTATACACAAAGTCTCAGCCTATCTCAACTGGATCTACATGT 1289
DB 1401 AGCTTACAGACAGGAGTGTACGGAAATGTGATGTTATTCAGGACTGGATTTATCGACA 1460
QY 1290 CTGGAAGGCTGA 1301
DB 1461 AATGAGGGCAGA 1472

RESULT 8
US-09-342-749-29
; Sequence 29, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtighian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/342,749
; EARLIER FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 60/091,044
; EARLIER FILING DATE: 1998-06-29
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-749-29

Query Match 11.3%; Score 147; DB 3; Length 2479;
Best Local Similarity 51.8%; Pred.No. 2.3e-31;
Matches 531; Conservative 0; Mismatches 460; Indels 38; Gaps 7;
QY 304 GTCCGCTCTCCAGGACCGATCCACACTGCGAGTGTGGACTCGGCCACAGGGAATCGG 363
DB 501 GTTCGCTCTACGAGCAAACTTCATCTTCAGATGTACTCATCTCAGAGGAAGTCTCTGG 560
QY 364 TTCTCTGCTGTTTCGACAACTTCAGAACTCTCGCTGAGACAGCTGTAGGAGATG 423
DB 561 CACCTCTGTGTCCAGACAGACTGGAACAGAACTACGGGCGGGGCTTCGAGGAGATG 620
QY 424 GGCTACAGCAGCAACCCACTTTTCAGAGTGTGGAGTTGGCCACAGCAGGATCTGGAT 483
DB 621 GGCTAAGATATTTTACTCTAG-CCAAGGATAGTGGATGACAGCGGATCCACAG 679
QY 484 GTTGTGAAATCAGAGAAACAGCAGGAGCTTCG-----CATGGGAATCAAG 533
DB 680 CTTTATGAACTGAACACAGAGTCCGCGCAATGTGATATCTATAAAAAACTGTACACAG 739
QY 534 TGGGCTGCTCTCTCAGGCTCCCTGCTCTCCCTGCTGCTCTGCTGCTGCTGCTGCTGCT 593

740 TGATGCTGTTCTTCAAAAGCAGTGGTTCTTTAGCGTGTAGCGCTCGGGGTCACATTT 799
594 GAAGACCCCGCC-----GTGTGGTGGGGAGGAGGCGCTCTGTGGATTCTTTGGCC 644
800 GAATCAAGCCCGCAGACAGATCGTGGCGGTGAGAGCGCGCTCCCGGGGCGCTGGCC 859
645 TTGGAGGTGACATCTCAGTACGACAAACAGACGCTCTGTGTGAGGAGGACATCTTGGACCC 704
860 CTGGAGGTGACCTGACGCTCCAGAGCTCCAGAGCTCCAGCTGTGGAGGCTCCATCATCACCCC 919
705 CCACTGGGTCTCAGCGAGCCACTGTCTCAGGAACAT-----ACCGATGTGTCA 757
920 CGAGTGGATCTGAGACCGCCCTCTGGTGGAAACCTCTTAACAATCCATGGCATTTG 979
758 ACTGGAAGGTGGCGGAGCTCAGACAAACTGGGAGCTTCCCATCCCTGCTGTGGCCA 817
980 GACGGCATTTGGCGGATTTTGGAGCAATCTTTCATGTTCTATGGAGCGGATACCAAGT 1039
818 AGATCATCATCATTAATTAACCCATGTACC-----CCAAAGACATGACATGGCCT 872
1040 ACAAAGATGATTTCTCATCCAAATATGACTCCAAAGCAAGAACAAATGATGATGGCT 1099
873 CATGAAGTGTGACCTTCTCAGGACAGTCCAGGCGCCATCTCTGCTGCGCCTT 932
1100 GATGAAGCTGCAAGAGCTCTGACTTTCAACGACCTAGTGAACCCAGTGTCTTGCACAA 1159
933 CTTTATGAGGAGCTCCTCAGGACCGCCCTCTGAGTCACTGTGGATCATTTGAGGCTTTACGAA 992
1160 CCGAGGATGATGCTGCGAGCCAGACAGCTCTGCTGGAATTTCCGGGTGGGGGCGCACCGA 1219
993 GCAGATGGAGGAAGATGCTGACATATCTGTGAGGCGTGTGAGGCGTGTGAGGCGTGTGAGG 1052
1220 GGAGAA---AGGAGAGCTCAGAGTGTGAACGCTGCGCAAGTCTTCTCATTTGAGAC 1276
1053 CACCGGTGCAATGACAGTGTGAGTGTGAGGAGGAGTGTGAGGAGGAGTGTGAGTGTG 1112
1277 ACAGAGATGCAACAGCAGATATGTCTATGACAAACCTGTATCACAACGAGCATGATCTGTGC 1336
1113 AGGCATCCCGAAGGGGTGTGACACCTGCGAGGTCACAGTGTGTGGGCGCCCTG---AT 1169
1337 CGGCTTCTGAGGGGAACGTGATTTTCCAGGAGTGTGAGGAGGCTCTGGTCA 1396
1170 GTACCAATCTGACAGTGGCATGTGGTGGCATCGTTAGCTGGGCTATGCTCGGGGG 1229
1397 TTGGAACAACAATATCTGGTGTGATAGGAGGATACAGAGTGGGCTTCTGGCTGTGCA 1456
1230 CCGAGACCCAGAGTATACACCAAGTGTGAGGCTCTGAGCTATCTCACTGGATCTACAATGT 1289
1457 AGCTTACAGACAGGAGTGTACGGAATGTGATGTATTCACGGACTGGATTTATCGACA 1516
1290 CTGGAAGGC 1298
1517 RATGAAGGC 1525

RESULT 9

US-09-691-840-29
; Sequence 29, Application US/09691840
; Patent No. 6444419
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavcigian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
; TITLE OF INVENTION: TSPRS2 is a Tumor Suppressor
; FILE REFERENCE: 2318-202
; CURRENT APPLICATION NUMBER: US/09/691,840
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US/09/342,749
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/091,044
; PRIOR FILING DATE: 1998-06-29

; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 29
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-691-840-29

Query Match

11.3%; Score 147; DB 4; Length 2479;
Best Local Similarity 51.6%; Pred. No. 2.3e-31;
Matches 531; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCGCGCTCTCAAGGACCGATCCACACTGCAAGGTGCTGAGTCTGGCCACAGGGAATGG 363
DB 501 GTTCGGCTCTACGGACAAACTTCTCCTTCAGATGACTCATCTCAGAGGAATCTCTGG 560
QY 364 TTCTGCTGCTGTTCGACAACTTCAGAAAGCTCTGTGAGACAGCTGTGAGGAGATG 423
DB 561 CACCTGTGTGCCAAGACGACTGGAACGAACTACGGCGGGCGGCTGCAAGGACATG 620
QY 424 GGCTACAGCAGCAAAACCCACTTTCAGAGCTGTGAGATTTGGCCACAGCAGATCTGGAT 483
DB 621 GGCTATAAGAAATATTTTACTCTAG-CCAAGAAATAGTGGATGACGGGATCCACAG 679
QY 484 GTTGTGAAATCACAGAAACACGACGAGGAGCTTCG-----CATGCGGAATCTCAAG 533
DB 680 CTTTATGAACATGAACACAAAGTCCGCAATGTCGATATCTATAAAATCTTACCACAG 739
QY 534 TGGGCGCTGTCTCAGGCTCCCTGTCTCCCTGACATCTTGTGCTGTGAGGAGCTT 593
DB 740 TGAATGCTGTCTTCAAAAGCAGTGTCTTCTTACGCTGTTTACGCTGCGGGGTCAACT 799
QY 594 GAAGACCCCGCC-----GTGTGGTGGGTGGGAGGAGGCTCTGTGATTTCTTGGCC 644
DB 800 GAATCAAGCCCGCAGACGAGGATCGTGGCGGTGAGAGCGCTCCCGGGGCGCTGGCC 859
QY 645 TTGGCAGGTGAGATCCAGTACGACAAACAGACAGTCTGTGAGGAGGAGATCTTGGACCC 704
DB 860 CTGGCAGGTGAGCTCAGCTCAGAACGCTCAGACGCTGCGGAGGCTCCATCATCACCCC 919
QY 705 CCACTGCTCTCAGCGAGCCACTGTCTCAGGAACAT-----ACCGATGTGTCA 757
DB 920 CGAGTGGATCTGAGAGCGCCCTCAGTGTGAGAAACCTCTTAACATCTCATGGCATG 979
QY 758 ACTGGAAGGTGGGCGAGGCTCAGACAAACTGGGAGCTTCCCATCCCTGCTGTGGCCA 817
DB 980 GACGGCATTTGGGGGATTTTGAGACAACTTTCATGTTCTATGGAGCGGATACCAAGT 1039
QY 818 AGATCATCATCATTAATTAACCCCTGTACC-----CCAAAGACATGACATCGCCT 872
DB 1040 ACAAAGTGTATTTCTATCCAAATATGACTCCAAAGCAAGATGAGATGATTTGGCT 1099
QY 873 CATGAAGTGTGAGTTCACACTCCTTCTCAGGACAGTCAAGGCGCCATCTGTGCTGCTT 932
DB 1100 GATGAAGCTGCAAGGCGCTGACTTTCAACGACCTAGTGAACCCAGTGTCTTGCACAA 1159
QY 933 CTTTATGAGGAGCTCCTCAGGACCGCCCTCTGAGTCACTGTGGATCATTTGAGGCTTTACGAA 992
DB 1160 CCGAGGATGATGCTGCGAGCAAGCAGCTCTGCTGGAATTTCCGGGTGGGGGCGCACCGA 1219
QY 993 GCAGATGGAGGAAGATGCTGACATATCTGTGAGGCGTGTGAGGCGTGTGAGGCGTGTGAGG 1052
DB 1220 GGAGAA---AGGAGAGCTCAGAGTGTGAACGCTGCGCAAGTCTTCTCATTTGAGAC 1276
QY 1053 CACCGGTGCAATGACAGTGTGAGTGTGAGGAGGAGTGTGAGGAGGAGTGTGAGTGTG 1112
DB 1277 ACAGAGATGCAACAGCAGATATGTCTATGACAAACCTGTATCACAACGAGCATGATCTGTGC 1336
QY 1113 AGGCATCCCGAAGGGGTGTGACACCTGCGAGGTCACAGTGTGTGGGCGCCCTG---AT 1169
DB 1337 CGGCTTCTGAGGGGAACGTGATTTTCCAGGAGTGTGAGGAGGCTCTGGTCA 1396
QY 1170 GTACCAATCTGACAGTGGCATGTGGTGGCATCGTTAGCTGGGCTATGCTCGGGGG 1229
DB 1397 TTGGAACAACAATATCTGGTGTGATAGGAGGATACAGAGTGGGCTTCTGGCTGTGCA 1456
QY 1230 CCGAGACCCAGAGTATACACCAAGTGTGAGGCTCTGAGCTATCTCACTGGATCTACAATGT 1289
DB 1457 AGCTTACAGACAGGAGTGTACGGAATGTGATGTATTCACGGACTGGATTTATCGACA 1516
QY 1290 CTGGAAGGC 1298
DB 1517 RATGAAGGC 1525

Db 1397 TTGAAACAAATATCTGGTGGCTGATAGGGGATACAGCTGGGGTCTTGCTCTGCGAA 1456
QY 1230 CCGAGACCCCGAGGATATACCAAGGCTCTCAGCCCTATCTCAACTGGATCTCAATGT 1289
Db 1457 AGCTTACAGACAGGAGTGTACGGGAATGTGATGTTTACGGAGCTGGATTTATCGACA 1516
QY 1290 CTGGAGGC 1298
Db 1517 AATGAAGC 1525

RESULT 10
US-09-685-166A-894
; Sequence 894, Application US/09685166A
; Patent No. 6630305
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriek
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 894
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-685-166A-894

Query Match 11.3%; Score 147; DB 4; Length 2479;
Best Local Similarity 51.6%; Fred. No. 2.3e-31;
Matches 531; Conservative 0; Mismatches 460; Indels 38; Gaps 7;

QY 304 GTCCGCCTCTCAAGGACCGATCCACACTGCGAGGTGCTGGACTCGGCCACAGGGAATGG 363
Db 501 GTTCGCTCTAGGACCAAACTTCATCCTTCAGATGTACTCATCTCAGAGGAAGTCCTGG 560
QY 364 TTCTCTGCTGTTTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCTGTAGGAGATG 423
Db 561 CACCCCTGTGTGCCAAGACGACTTGAACAGAACTACGGCGGGCGGCTGTGAGGACATG 620
QY 424 GGCTACAGCAGCAACCCACTTTTCAGAGCTGTGGAGATTGCCACAGCAGGATCTGAT 483
Db 621 GGCTATAGATATATTTTACTCTAG-CCAAGATAGTGTATGATCAGCGGATCCACAG 679
QY 484 GTTGTGAAATCACAAGAAACAGCCAGGAGTTGG- - - - -CATGCGGAATCAAG 533
Db 680 CTTTATGAACTGAACACAGAGTGGCGCAATGTCGATATCTATAAAAACTGTACACAG 739
QY 534 TGGGCGCTGTCTCTCAGGCTCCCTGCTCCCTGCACTGTCTGCTGTGGGAGAGCT 593
Db 740 TGATCGCTGTCTCTCAAAGACAGTGGTTCTTACGCTGTAGCTTGGGCGGCTCAATT 799
QY 594 CAAAGACCCCCC- - - - -GTGTGGTGGGTGGGAGGAGGCTCTGTGGATTTCTGGCC 644

Db 800 GAACTCAAGCCGCCAGACAGGATCTGTGGCGGTGAGAGCGGCTCCCGGGGCGCTGGCC 859
QY 645 TTGCGAGGTGAGCATCCAGTACAGAAACAGCAGCTCTGTGAGGAGAGATCTGTGACCC 704
Db 860 CTGGCAGGTGAGCTGACAGTCCAGAAAGTCCAGGTGTGCGGAGGCTCCATCATCACCC 919
QY 705 CCACTGGGTCTTCAAGGAGCCAGCTCTTCAGGAACAT- - - - -ACCGATGTCTCA 757
Db 920 CGAGTGTGATCGTACAGCCCGCCCTGCGTGGAAACCTTTTAAACAATCCATGGCATTTG 979
QY 758 ACTGGAAGGTGCGGCGAGGCTCAGACAAACTGGGCGAGCTTCCCATCCCTGGCTGTGGCCA 817
Db 980 GACGGCATTTGCGGGATTTTGAGACAATCTTTTCATGTTCTATGGAGCCGATACCAAGT 1039
QY 818 AGATCATCATATTGAATTCACCCCATGTACC- - - - -CCAAAGACAATGACATCGCCCT 872
Db 1040 AAAAAAGTGATTTCTCATCCAAATTTGACTCCAAAGACCAAGAACAAATGACATTCGCT 1099
QY 873 CATGAAGCTGCAAGTTCCTCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTT 932
Db 1100 GATGAAGCTGCAGAGCTCTGACTTTCAACGACCTAGTGAACACAGTGTGTCTGCCCA 1159
QY 933 CTTTGTAGGAGCTCACTCCAGCCACCCCACTCTGTGATCACTTGGATGGGGCTTTACGAA 992
Db 1160 CCGAGGATGATGCTGCGAGCCAGACAGCTCTGCTGGATTTCCGGGTGGGGGCGCCACCGA 1219
QY 993 GCAGATGGAGGGAAGATGTCTGACATATCTGTGAGGCGTCACTCCAGTCACTTACACAG 1052
Db 1220 GGAGAA- - - - -AGGGAAGACCTCAGAACTGTGAAGCTGCAAGTCTTCTCATTCAGAC 1276
QY 1053 CACAGGTGCAATGCAGACGATCGTACCAGGGGGAAGTCAACGAGAAGATGATGTGTC 1112
Db 1277 ACAGAGATGCAACAGCAGATATGTCTATGACAACTGATCACACACAGCATGATGTGTC 1336
QY 1113 AGCATCCCGGAAGGGGTGTGACACCTCCAGGCTGACAGTGTGGGCGGCTG- - -AT 1169
Db 1337 CGGCTTCTGCGAGGGAACGTGATTTCTGCCAGGTGACAGTGGGGGCTCTGTCTAC 1396
QY 1170 GTACCAATCTGACCAGTGGCATGTGTGGGCGATCGTTAGCTGGGGCTATGGCTGCGGGG 1229
Db 1397 TTCGAACAACAATATCTGGTGGCTGATAGGGGATACAGCTGGGGTCTGGCTGTGCCAA 1456
QY 1230 CCGAGACCCCGAGGATATACACAGGCTCTCAGGCTATCTCAACTGATCTACATGT 1289
Db 1457 AGCTTACAGACAGGAGTGTACGGGAATGTGATGTTTACCGGACTGGATTTATCGACA 1516
QY 1290 CTGGAGGC 1298
Db 1517 AATGAAGC 1525

RESULT 11
US-08-807-151-2
; Sequence 2, Application US/08807151
; Patent No. 6043033
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; TITLE OF INVENTION: PROTEASE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/807,151
 ; FILING DATE: Filed Herewith
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0227 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1077 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: SCORN0T01
 ; CLONE: 556016
 ; US-08-807-151-2

Query Match 10.9%; Score 142.8; DB 3; Length 1077;
 Best Local Similarity 53.3%; Pred. No. 2.4e-30;
 Matches 428; Conservative 0; Mismatches 348; Indels 27; Gaps 5;

QY	523	CGGAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGGCTGT	582
Db	146	CTGTACCACAGTATGCTCTTCTCAAGAGAGTGGTCTTACGCTGTATAGCTGC	205
QY	583	GGGAAGAGCTGAAGACCCCTCC-----GTGTGGTGGTGGGAGGAGGCTTGTG	633
Db	206	GGGGTCAACTTGAACCTCAAGCGCGCAGACAGGATCGTGGGCGGAGGCGCTCCCG	265
QY	634	GATTTCTGGCTTGGCAGTGCAGTACAGTACGACAAACAGCAGCTGTGGAGGAGC	693
Db	266	GGGGCTGGCTGGCAGTGCAGTACGCTGACGCTCAGACGCTCCAGCTGTGGAGGCTCC	325
QY	694	ATCTGGACCCCACTGGGTCTCAGGAGCCCACTGCTCAGGAACAT-----AC	746
Db	326	ATCATCACCCCGAGTGGATCGTACAGCGGCCCACTCGTGGGAAACCTCTTAACAAT	385
QY	747	CGATGTGTTCAACTGGAAGTGGGCGAGGCTCAGACAACTGGGCGCTTCCATCCT	806
Db	386	CCATGGCATGGACGGCATTTGCGGGGATTTGAGACAACTCTTTCATGTTCTATGGAGCC	445
QY	807	GGCT-----GTGGCAAGATCATCATTAATTAACCCCACTGTACCCCAAGACAAT	861
Db	446	GGATACCAAGTAGAAGAGTATTTCTCATCCAAATTATGACTCCAGACCAAGACAAT	505
QY	862	GACATGCGCTCATGAAGTGGCTTCCCACTCACTTTCTCAGCAGTACAGCCCATC	921
Db	506	GACATGCGCTCATGAAGTGGCAGAGCTTCACTTTCAACGACCTAGTGAACACCAATG	565
QY	922	TGCTGCGCTCTTTTGTATGAGAGTCACTCCAGCCACCCCACTGTGATCATTTGGATGG	981
Db	566	TGCTGCGCTCAACCCAGGATGATGTGCGCCAGACAGCTCTGCTGATTTCCGGGTGG	625
QY	982	GGCTTTAGACAGCAAGTGGAGGAGATGTGTGACATCTGCTGCGAGGCTGATGTCAG	1041
Db	626	GGGGCCACCGAGGAGAA---AGGGAGACCTCAGAAAGTGTGAACGCTGCGCAAGGTGCTT	682
QY	1042	GTCATTGACAGCACAGGTGCAATCCAGACATGCTTACCGGGGGAAGTCAACGAGAAG	1101
Db	683	CTCATTTGACACACAGATGACACAGCAGATATGTCTATGACACCTGATCACACAGCC	742
QY	1102	ATGATGTGTGAGGATCCCGGAAGGGGTGTGACACTCCAGCGGTGTGACAGTGTGGG	1161
Db	743	ATGATCTGTGCGCGTCTCTGAGGGGAACGTGATTTCTTCCAGGGGTGACAGTGGAGG	802

QY 1162 CCCTG---ATGTACCAATCTGACAGTGGCATGTGGTGGCATCTGTAGCTGGGGCTAT 1218
 Db 803 CNTCTGGTCACTTCGAAGAACAATATCTGTGGCTGATAGGGGATACAAGCTGGGGTCT 862
 QY 1219 GGCTGGGGGGCGGAGCACCCCGAGGATATACCAAGGTCTCAGCCCTATCTCAACTGG 1278
 Db 863 GGCTGTGCCAAGCTTACAGACCGAGGTGTACGGGAATGTGATGTTACGGGACTGG 922
 QY 1279 ATCTACAATGTCTGGAAGCTGA 1301
 Db 923 ATTTATCGAACAATGAGGGCAGA 945

RESULT 12
 US-09-478-957-2
 ; Sequence 2, Application US/09478957
 ; Patent No. 6350448
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Lal, Preeti
 ; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
 ; TITLE OF INVENTION: PROTEASE
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/478,957
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA: 08/807,151
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0227 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1077 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: SCORN0T01
 ; CLONE: 556016
 ; US-09-478-957-2

Query Match 10.9%; Score 142.8; DB 4; Length 1077;
 Best Local Similarity 53.3%; Pred. No. 2.4e-30;
 Matches 428; Conservative 0; Mismatches 348; Indels 27; Gaps 5;

QY	523	CGGAACCTCAAGTGGGCGCTCTCTCAGGCTCCCTGGTCTCCCTGCACTGCTTGGCTGT	582
Db	146	CTGTACCACAGTATGCTCTTCTCAAGAGAGTGGTCTTACGCTGTATAGCTGC	205
QY	583	GGGAAGAGCTGAAGACCCCTCC-----GTGTGGTGGTGGGAGGAGGCTTGTG	633
Db	206	GGGGTCAACTTGAACCTCAAGCGCCAGACAGATGCTGGGCGGAGAGCGGCTCCCG	265
QY	634	GATTTCTGGCTTGGCAGGCTCAGCATCCAGTACGACAAACAGCAGCTCTGTGGAGGAGC	693

Db 266 GGGGCTGGCCCTGGCAGGCTCAGCCTGCACGTCAGAACGTCACGCTGGCGAGGCTCC 325
Qy 694 ATCTGGAGCCGCCCTGGCTCTCACGGCAGCCCACTGCTTCAGGAAACAT-----AC 746
Db 326 ATCATCACCCTCCGAGTGGATGTCAGACAGCGCCCACTGGGAAACCTTTAAACAT 385
Qy 747 CGATGTTCACTGGAAGTGGGGCAGGCTCAGACAACTGGGCGAGCTTCCATCCCT 806
Db 386 CCATGGCATGGACGGCATTTGGCGGGATTTTGAGACAATCTTTCACTGTTCTATGGAGCC 445
Qy 807 GGCT-----GTGGCCAGAGATCATCATTTGAATTTCAACCCCATGTACCCCAAGACAAT 861
Db 446 GSATACCAAGTAGAAAAGTGAATTCATCAATCAATTAATGACTCCAGACCAAGACAAT 505
Qy 862 GACATGCGCCCTCATGAAGTGCAGTTCCTCACTACTTCTCAGGCAAGTCAGGCCCCATC 921
Db 506 GACATTCGGCTGATGAAGCTCAGAACCTCTGACTTTCAACGACCTAGTGAACCAAGTG 565
Qy 922 TGTCTGCCCTCTTTGATGAGGAGCTCACTCCAGCCACCCCACTCTCGATCATTTGGATGG 981
Db 566 TGTCTGCCCAACCCAGGATGATCTGCAGCCAGAACAGCTCTCTGATTTCCGGTGG 625
Qy 982 GCTTTACGAAGCAGATGGAGGAAAGTGTCTGACATATCTGCTGAGGGCTCAGTCCAG 1041
Db 626 GGGGCCACCGAGGAGAA---AGGGAAGACCTCAGAGTGTGTAACGCTGCCAAGTCTT 682
Qy 1042 GTCTATTGACACACACGCTGCAATGTCAGACGATCGTACAGGGGGAAGTCACCGAGAAG 1101
Db 683 CTCATTGAGACACAGAGATGCAACAGCAGATATCTCTATGACACCTGATCACAACGACC 742
Qy 1102 ATGATGTGTGAGGATCCCGGAAGGGGTGTGACACACTCCAGGGGTGACAGTGGTGGG 1161
Db 743 ATGATCTGCGCGCTTCTGCAGGGGAACGTCGATTTGCGCAGGCTGACAGTGGAGGG 802
Qy 1162 CCCCCTG---ATGTACCAATCTGACAGTGGCATGTGTGGGCACTGTAGTGGGGCTAT 1218
Db 803 CNTCTGTCTCACTTCGAAGAACATATCTGTGTGGTGTATAGGGGATACAAGCTGGGGTCT 862
Qy 1219 GGCTGGGGGGCCGAGCACCAGGAGTATACACCAAGGCTCTCAGCCTATCTCACTGG 1278
Db 863 GGCTGTGCAAGCTTACAGACCAAGAGTGTACGGGATGTATGATGATTCACGGACTGG 922
Qy 1279 ATCTACCAATGTCTGGAAGGCTGA 1301
Db 923 ATTTATCGACAATGAGGGCAGA 945

RESULT 13

US-09-261-416-1
; Sequence 1, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 1
; LENGTH: 2416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: 144..1511
; OTHER INFORMATION: CDS

US-09-261-416-1

Query Match 10.2%; Score 133.6; DB 3; Length 2416;
Best Local Similarity 53.4%; Pred. No. 1.3e-27;
Matches 421; Conservative 0; Mismatches 349; Indels 18; Gaps 6;

RESULT 14

US-09-518-046-3
; Sequence 3, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518.046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261.416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153

Qy 532 AGTGGGCCCTGTCTCTCAGGCTCCCTGTGCTCCCTGCACCTGCTTGCCTGTGGAGAGAC 591
Db 717 AGGGAGGGATGTGCCTCTTGCCACCGTGTACCTTTGCACTGCACAGCCCTGTGTATAGA 776
Qy 592 CTG---AAGACCCCTCGTGTGGTGGGAGGAGGCTCTGTGATTTCTTGGCCTTGG 648
Db 777 AGGGGCTACAGCTCAGCATCGTGGTGGAAACATGTCTTCTCTCAGAGTGGCCCTGG 836
Qy 649 CAGGTTCAGCATCCAGTACGACAAACAGACAGCTGTGTGGAGGAGGATCTCTGACCCCAAC 708
Db 837 CAGGCCAGGCTTCAGTTCAGGCTTACCACTGTGCGGGGCTCTGTCTATCAGCCCTG 896
Qy 709 TGGGTCTTCAGGCGCCCACTGCTT-----CAGGAACATACGATGTGTCACTGG 762
Db 897 TGGATCATCAGCTGCTGACACATGTGTATGACTTGTACCTCCCAAGTCACTGAGCAATC 956
Qy 763 AAGGTGCGGGCAGGCTCAGACAACTGGGAGCTTCCCATCTCCCTGGCTGTGGCCCAAGATC 822
Db 957 CAGGTGGGTCTAGTTTCCCTGTGGACAAATCCAGCCCCCACTCCACTTGTGGAGAGATT 1016
Qy 823 ATCATCATTTGAATTAACCCCATGTACCCCAAGACAAATGACATGCCCTCATGAGCTG 882
Db 1017 GTTTACCAACAGCAAGTACAAAGAGGCTGGGCAATGACATGCCCTTATGAAGCTG 1076
Qy 883 CAGTTCCCACTCACTTTCTCAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAG 942
Db 1077 GCGGGCGCACTCAGTTCAATGAATGATCCAGCTGTGTGCTGCCCACTCTGAGAG 1136
Qy 943 GAGCTCACTCCAGCACCCCACTCTGGATCATTTGATGGGGCTTTTACGAAGCAGATGGA 1002
Db 1137 AACTTCCCGATGGAAGTGTGCTGAGCTCAGGATGGG---GGCCACAGAGGATGGA 1193
Qy 1003 GGGAGATGCTGACATCTGTGAGGGGTGAGTCCAGCTCATTTGACAGCACA-CGGTG 1061
Db 1194 GGTGACGCTCCCTGTCTGACACCGCGGCTTCTTGTATTTCCAAAGATCTG 1253
Qy 1062 CAATGACAGCATGTCTACAGGGGAAGTCCACGAGAAAGATGATGTGTGCAAGGCATCCC 1121
Db 1254 CAACACAGAGGAGCTGTACGGTGGCATCATCTCCCTCCATGCTCTGCGCGGCTTACCT 1313
Qy 1122 GGAAGGGGTGT---GGACACCTGCGAGGCTGACAGTGTGGGGCCCTGATGTACCAATCT 1179
Db 1314 GACGGTGGCGTTGGAAACAGCTGCCAGGGGACAGCGGGGGGCTGTGTGTCAAGAG 1373
Qy 1180 GAC---CAGTGGCATGTGTGGGCACTGTAGTGGGGCTATGGCTGCGGGGGCCCGAGC 1236
Db 1374 AGGAGCTGTGGAAGTTAGTGGAGCGCAGCTTTGGCATCGGCTGCCAGACGTGAAC 1433
Qy 1237 ACCCAGGAGTATACCAAGGCTCTCAGCCTATCTCAACTGGATCTACAAATGTCTGGAAG 1296
Db 1434 AAGCCTGGGCTGTACACCGGTGTCACTCTCTCTGGACTGGATCCACGAGCAGATGGAG 1493
Qy 1297 GCTGAGCT 1304
Db 1494 AGAGACCT 1501

RESULT 14

US-09-518-046-3
; Sequence 3, Application US/09518046
; Patent No. 6294663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
; TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
; FILE REFERENCE: D6192CIP
; CURRENT APPLICATION NUMBER: US/09/518.046
; CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261.416
; EARLIER FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 153

SEQ ID NO 3
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
OTHER INFORMATION: entire cDNA sequence of TADG-12 variant gene
US-09-518-046-3

Query Match 9.4%; Score 122.8; DB 3; Length 2544;
Best Local Similarity 55.3%; Pred. No. 1.5e-24;
Matches 282; Conservative 0; Mismatches 22; Indels 6; Gaps 2;
QY 798 CCATCCCTCGCTGTGGCAAGATCATCATCTGAATTCACCCCAAGTACCCCAAGA 857
DB 1123 CCATCCCACTTGGTGGAGAGATTGTTACACAGCAAGTACAAGCCAAAGAGCGCTGG 1182
QY 858 CAATGACATCGCCCTCATGACGCTGAGTCCCACTCTTCTCAGCACAGTCAGGCC 917
DB 1183 CAATGACATCGCCCTCATGACGCTGCGCGGCCACTCAGCTTCAATGAATGATCCAGCC 1242
QY 918 CATCTGTCTGCCCTTCTTTGATGAGGAGCTCACTCCAGCCACCCCACTCTTGGATCATGG 977
DB 1243 TGTGTGCTGCCCACTCTGAAGAACTTCCCGATGAAAGTGTGCTGGACGTCAGG 1302
QY 978 ATGGGGCTTTAGCAAGCAGATGAGAGGAAGATGTCTGACATCTGCTGAGGGGTCACT 1037
DB 1303 ATGGGG--GGCCACAGAGGATGGAGGTGACGCTCCCTCTGTGAACCAACGCGGCGGT 1359
QY 1038 CCAGGTCTATTGACAGACACAGGTGCAATGACAGACGATCGTACCAGGGGGAAGTCAACGA 1097
DB 1360 CCTTTGATTCCACAGATCTGCAACACAGGACGTGTACGGTGCATCATCTCCCC 1419
QY 1098 GAAGATGATGTGACGATCCCGAAGGGGTGTGACACCTGCCAGGTGACAGTGG 1157
DB 1420 CTCATGCTCTCGCGGGCTACTCTGACGGTGGCGTGGACAGCTGCCAGGGGACAGCGG 1479
QY 1158 TGGGCCCTGATGATACCATCTGAC--CAGTGGATGTGTGGGCATCGTTAGCTGGG 1214
DB 1480 GGGGCCCTGTGTGTCAAGAGAGAGGCTGTGGAGGTGTAGTGGAGCGACCACTTTGG 1539
QY 1215 CTATGGCTGCGGGGCGCGAGCACCCAGGAGTATACCAAGGTCTCAGCCTATCTCAA 1274
DB 1540 CATCGCTGCGCAGAGGTGAACAACCTGGGTGTACACCCGTGTACACCTCTCTCTCTGGA 1599
QY 1275 CTGGATCTACATGTCTGGAAGGCTGACCT 1304
DB 1600 CTGGATCCACGACGATGCGAGAGACCT 1629

RESULT 15
US-09-636-382A-3
Sequence 3, Application US/09636382A
Patent No. 6514741
GENERAL INFORMATION:
APPLICANT: Presnell, Scott R.
APPLICANT: Taft, David W.
TITLE OF INVENTION: TRYPTASE-LIKE POLYPEPTIDE ZTRYP1
FILE REFERENCE: 99-21
CURRENT APPLICATION NUMBER: US/09/636,382A
CURRENT FILING DATE: 2000-08-09
PRIOR APPLICATION NUMBER: US 60/149,563
PRIOR FILING DATE: 1999-08-18
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 942
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate polynucleotide sequence of mouse ztryp1
OTHER INFORMATION: (SEQ ID NO:2)
NAME/KEY: misc_feature

LOCATION: (1)...(942)
OTHER INFORMATION: n = A,T,C or G
US-09-636-382A-3

Query Match 9.1%; Score 119.4; DB 4; Length 942;
Best Local Similarity 33.8%; Pred. No. 8.4e-24;
Matches 247; Conservative 122; Mismatches 337; Indels 24; Gaps 4;
QY 578 CTTGTGGGAAGAGCCCTGAAGAGCCCCCGTGTGGTGGGGAGGAGGCGCTCTGTGGATT 637
DB 101 SNTGYGNCARACNAAYATHACNTGYAARGTNGTNAAYGNAARGCNGTNGARGTNGNA 160
QY 638 CTTGGCTTGGCAGGTGACGATCCAGTACGACAAACAGCACGCTCTGTGAGGAGCATCC 697
DB 161 ARTGGCCTGGCARGTNGSNATHYNTTYYTNGNATGTAYATHGTGYWSNGNWSYNTNA 220
QY 698 TGGACCCCACTGGGTCTCAGGCGAGCCCACTGCTTCAAGAAACATACCGATGTGTCA 757
DB 221 THCAYCAYCAYTGGATHYTNACNGCNCNCAVTGYTNCARMGNWSNARAAYCCGNA 280
QY 758 ACTGGAAGTGGGGCAGGCTCAGACAAACT---GGGCGAGTTCCTCCATCCCTGGCTGG 814
DB 281 ARTAYACNGTNAARGTNGGNTNCAPACNYTCCNGAYAAAYWSNACNWSNGARYNTYNG 340
QY 815 CCAAGATCATCATCTGAATTCAACCCCACTGTACCCCAAGACATGACATCGCCTCA 874
DB 341 TNAONMGNATHGNTNTHCAYGARAATYTTAAAYMGNATGWSNGAYGAYATHGNAHY 400
QY 875 TGAAGCTGCAAGTTCCTCACTCCTTTCTCAGGCACAGTCAAGGCCCATCTGTCTGCCCTTCT 934
DB 401 TNAARYTNAARTAYCCNGTNACTNGWSNCCNYTNGTNCARCNCNATHIGYTYNCCNWSNT 460
QY 935 TTGATCAGAGGACTCCTCCAGCCACCCCACTCTGGATCATTTGGATGGGCTTTACGAGC 994
DB 461 TYAAYTNAARCCNWSNATHGNGNACNATGTGTGGTNGTNGTNGGNGYNGARAARG 520
QY 995 AGAATGGA---GGGAAGATGTCTGACATCTGTGCGAGCGCTCAGTCCAGGTCAATTGACA 1051
DB 521 CNGARGNCAYCCNARAACNCCNTAYWSNGTNCARGNYTNGCNGTNGMGNATHGTAAYA 580
QY 1052 GCACAGGTGCAATGACAGACGATGCGT-----ACCAGGGGGAAGTCAACCG 1096
DB 581 AYGARATHGTAAAYCAYMGNATYCATYTYTNTYNTYNTYNTYNTYNTYNTYNTYNTYNTY 640
QY 1097 AGAAGATGATGTGCGAGCATCCCGGAAGGGGTGTGACACACCTGCCAGGTGACAGTG 1156
DB 641 GNDAYGAYATGYNTGYACNWSNWSNGARTGGGNYTNGAYACNTGYCARGAYACNWSNG 700
QY 1157 GTGGGCCCTGATGTATACCAATCTGACAG---TGGCATGTGTGGGCAATCGTGTAGCTGG 1213
DB 701 GNWSNWSNTNGTNTGYCARATGAAYARACNTGGGTNCARATGGGNGTNGTNGTNGTNG 760
QY 1214 GCTATGCTGCGGGGCGCGAGCACCCAGGAGTATACACCAAGGTCTCAGCCTATCTCA 1273
DB 761 AYTTCATGTGNGMGNMGNCACTYCCNWSNGTNTAYACNWSNACNWSNCACTTYACNC 820
QY 1274 ACTGGATCTA 1283
DB 821 ARTGGATHAA 830

Search completed: June 8, 2004, 02:22:27
Job time : 123 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2004, 01:16:00 ; Search time 598 Seconds
(without alignments)
9955.526 Million cell updates/sec

Title: US-10-030-688-1

Perfect score: 1305

Sequence: 1 atggatctgacagtgatca.....atgctggaaggtgagctg 1305

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2995936 seqs, 2280998010 residues

Total number of hits satisfying chosen parameters: 5991872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303.4	99.9	2038	13	US-09-968-415-18
2	1303.4	99.9	2038	15	US-10-180-719-18
3	1303	99.8	1314	16	US-10-295-027-133
4	1303	99.8	1314	16	US-10-295-027-778
5	1303	99.8	1314	16	US-10-295-027-790
6	1303	99.8	1314	16	US-10-295-027-830
7	1303	99.8	1314	16	US-10-295-027-979
8	1303	99.8	1314	16	US-10-173-399-88
9	1303	99.8	2081	9	US-09-851-588-7
10	1303	99.8	2307	15	US-10-097-340-317
11	1303	99.8	2307	15	US-10-171-311-217
12	1298.2	99.5	2137	10	US-09-776-191-3
13	1298.2	99.5	2137	15	US-10-156-214B-3
14	1294	99.2	2165	15	US-10-101-510-634

```
15 1292 99.0 2079 9 US-09-851-588-5
16 1292 99.0 2079 10 US-09-776-191-71
17 1292 99.0 2079 15 US-10-264-820-22
18 1292 99.0 2079 15 US-10-254-289-1
19 1292 99.0 2079 16 US-10-156-214A-38
20 1263 96.8 2063 10 US-09-888-257A-2
21 1263 96.8 2063 10 US-09-946-374-274
22 1263 96.8 2063 12 US-10-015-395A-274
23 1263 96.8 2063 13 US-10-206-915-329
24 1263 96.8 2063 13 US-10-199-670-329
25 1263 96.8 2063 13 US-10-201-858-329
26 1263 96.8 2063 13 US-10-205-890-329
27 1263 96.8 2063 13 US-10-208-024-329
28 1263 96.8 2063 13 US-10-201-853-329
29 1263 96.8 2063 13 US-10-063-745-111
30 1263 96.8 2063 13 US-10-063-512-111
31 1263 96.8 2063 13 US-10-063-513-111
32 1263 96.8 2063 13 US-10-063-549-111
33 1263 96.8 2063 13 US-10-063-589-111
34 1263 96.8 2063 13 US-10-063-551-111
35 1263 96.8 2063 13 US-10-174-581-329
36 1263 96.8 2063 13 US-10-176-483-329
37 1263 96.8 2063 13 US-10-176-749-329
38 1263 96.8 2063 13 US-10-176-914-329
39 1263 96.8 2063 13 US-10-176-915-329
40 1263 96.8 2063 13 US-10-006-485A-274
41 1263 96.8 2063 13 US-10-013-907A-274
42 1263 96.8 2063 13 US-10-015-499A-274
43 1263 96.8 2063 13 US-10-063-555-111
44 1263 96.8 2063 13 US-10-063-563-111
45 1263 96.8 2063 13 US-10-063-594-111
```

ALIGNMENTS

RESULT 1

US-09-968-415-18
Sequence 18, Application US/09968415
Publication No. US20020086334A1

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/968,415
FILING DATE: 26-Sep-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/659,151

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

```

, , TELEPHONE: 650-855-0555
, , TELEFAX: 650-845-4166
, , INFORMATION FOR SEQ ID NO: 18:
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 2038 base pairs
, , TYPE: nucleic acid
, , STRANDEDNESS: single
, , TOPOLOGY: linear
, , IMMEDIATE SOURCE:
, , LIBRARY: COLNNOT13
, , CLONE: l337018
, , SEQUENCE DESCRIPTION: SEQ ID
US-09-968-415-18

```

Query Match	99.9%	Score	1303.4	DB	13	Length	2038
Best Local Similarity	99.9%	Pred. No. 0					
Matches 1304	Conservative	0	Mismatches	1.1	Indels	0	Gaps
QY	1	ATGATCTGACAGCTGATCAACCTCTGAAACAGCCTCGATGTCAAAACCCCTCGCAAAACCC	60				
DB	200	ATGATCTGACAGCTGATCAACCTCTGAAACAGCCTCGATGTCAAAACCCCTCGCAAAACCC	259				
QY	61	CGATGCCCATGAGAGCCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC	120				
DB	260	CGATGCCCATGAGAGCCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGC	319				
QY	121	CTGGCGAGTATCATATGTTGGTTCCTCATCCAGGTGATTCCTGGATAAATACTACTTC	180				
DB	320	CTGGCGAGTATCATATGTTGGTTCCTCATCCAGGTGATTCCTGGATAAATACTACTTC	379				
QY	181	CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAAGACAGCTGTGTGACGGAGAGCTGGAC	240				
DB	380	CTCTCGGGGAGCCTCTCCACTTCATCCGAGGAAGACAGCTGTGTGACGGAGAGCTGGAC	439				
QY	241	TGTCCTTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAATG	300				
DB	440	TGTCCTTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCCTGCAATG	499				
QY	301	GCAGTCCGCTCTCTCAAGGACCGATCCACACTGCAGGTCTGGACTCGGCCACACGGGAAC	360				
DB	500	GCAGTCCGCTCTCTCAAGGACCGATCCACACTGCAGGTCTGGACTCGGCCACACGGGAAC	559				
QY	361	TGTTCTCTGCTGTTTCGACAACTTCAAGAGCTCTCGTGAAGACAGCCTGTAGGAG	420				
DB	560	TGTTCTCTGCTGTTTCGACAACTTCAAGAGCTCTCGTGAAGACAGCCTGTAGGAG	619				
QY	421	ATGGCTACAGCAACCAACCTTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTG	480				
DB	620	ATGGCTACAGCAACCAACCTTTTCAGAGCTGTGGAGATTGGCCACAGACAGGATCTG	679				
QY	481	GATGTGTGAAATCAAGAAAAACGCCAGGAGCTTCGCATGCGGAAATCAAGTGGGCC	540				
DB	680	GATGTGTGAAATCAAGAAAAACGCCAGGAGCTTCGCATGCGGAAATCAAGTGGGCC	739				
QY	541	TGTCCTTCAGGCTCCCTGGTCTCCCTGCACCTGCTGCTGTGGAGAGCCTCAAGACC	600				
DB	740	TGTCCTTCAGGCTCCCTGGTCTCCCTGCACCTGCTGCTGTGGAGAGCCTCAAGACC	799				
QY	601	CCCGTGTGGTGGGGAGGAGGCTCTGTGGATTCTTGGCCCTTGGCAGGTCAAGCATC	660				
DB	800	CCCGTGTGGTGGGGAGGAGGCTCTGTGGATTCTTGGCCCTTGGCAGGTCAAGCATC	859				
QY	661	CAGTACGACAAACAGCAGTCTGTGGAGGAGCATCTGGACCCCACTGGGGTCTCAAG	720				
DB	860	CAGTACGACAAACAGCAGTCTGTGGAGGAGCATCTGGACCCCACTGGGGTCTCTCAAG	919				
QY	721	CGAGCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGGGGCAGGCTCA	780				
DB	920	CGAGCCCACTGCTTCAGAAACATACCGATGTGTTCAACTGGAAGGTGGGGCAGGCTCA	979				
QY	781	GACAAACGTGGGACGCTCCCATCCCTGGCTGTGGCAGATCATCATATTGAATTCAC	840				
DB	980	GACAAACGTGGGACGCTCCCATCCCTGGCTGTGGCAGATCATCATATTGAATTCAC	1039				

RESULT 2

US-10-180-719-18
Sequence 18, Application US/10180719
Publication No. US20030166246A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Hillman, Jennifer L.
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSO for Windows Version 2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/180,719
FILING DATE: 25-JUN-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271
FILING DATE: 16-Jan-1998
ATTORNEY/AGENT INFORMATION:
NAME: Moran-Peterson, Sheela
REGISTRATION NUMBER: 41,201
REFERENCE/DOCKET NUMBER: PF-0458 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-180-719-18

Query Match 99.9%; Score 1303.4; DB 15; Length 2038;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCATCAACCCCTGCGCAACCC 60
DB 200 ATGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCATCAACCCCTGCGCAACCC 259

QY 61 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 260 CGTATCCCATGGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 319

QY 121 CTGCGAGTATCATCATTTGTTTCTCTCATCAAGTGTATTCGGATAAATACTACTTC 180
DB 320 CTGCGAGTATCATCATTTGTTTCTCTCATCAAGTGTATTCGGATAAATACTACTTC 379

QY 181 CTCTGGGGGACCTCTCCATTCATCCGAGGAGCAGCTGTGTGAGGAGCTGGAC 240
DB 380 CTCTGGGGGACCTCTCCATTCATCCGAGGAGCAGCTGTGTGAGGAGCTGGAC 439

QY 241 TGTCCCTTTGGGGAGGAGCAGGAGCAGCTGTGTGAGGAGCTTCCCGAAGGCGCTGCAGTG 300
DB 440 TGTCCCTTTGGGGAGGAGCAGGAGCAGCTGTGTGAGGAGCTTCCCGAAGGCGCTGCAGTG 499

QY 301 CGAGTCCGCTCTCCAGGAGCGATCCACACTGACAGTGTGTGAGTGTGAGGAGCAGGAGC 360
DB 500 CGAGTCCGCTCTCCAGGAGCGATCCACACTGACAGTGTGTGAGTGTGAGGAGCAGGAGC 559

QY 361 TGGTCTCTGCTGTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCCTGTAGGAGCAG 420
DB 560 TGGTCTCTGCTGTTTCGACAACTTCACAGAGCTCTGCTGAGACAGCCTGTAGGAGCAG 619

QY 421 ATGGGCTACAGCAGCAACCCACTTTTACAGCTGTGAGATGTGAGGAGTGTGAGGAGCAGG 480
DB 620 ATGGGCTACAGCAGCAACCCACTTTTACAGCTGTGAGATGTGAGGAGTGTGAGGAGCAGG 679

QY 481 GATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGTGGGCCC 540
DB 680 GATGTTGTTGAATCACAGAAACAGCCAGGAGCTTCGATGCGGAACCTCAAGTGGGCCC 739

QY 541 TGTCTCTCAGGCTCCCTGCTCTCCCTGACACTGTCTTGGCTGTGGAAGAGCCTCAAGAC 600
DB 740 TGTCTCTCAGGCTCCCTGCTCTCCCTGACACTGTCTTGGCTGTGGAAGAGCCTCAAGAC 799

QY 601 CCCGCTGTGTTGGTGGGAGAGGCGCTCTGTGATTTTGGCTTGGCTTGGCAGTCAAGATC 660
DB 800 CCCGCTGTGTTGGTGGGAGAGGCGCTCTGTGATTTTGGCTTGGCAGTCAAGATC 859

QY 661 CAGTACGCAAAACAGCAGCTGTGAGGAGGAGCCTCTGGAACCCCTCAAGTGGCTCTCAGC 720
DB 860 CAGTACGCAAAACAGCAGCTGTGAGGAGGAGCCTCTGGAACCCCTCAAGTGGCTCTCAGC 919

QY 721 GCAGCCCACTGTTTCAAGAAACATACCGATGTGTTCAACTGGAAGGTCGGGCGAGGTC 780
DB 920 GCAGCCCACTGTTTCAAGAAACATACCGATGTGTTCAACTGGAAGGTCGGGCGAGGTC 979

QY 781 GACAACTGGGAGCTTCCCATCCCTGCTGTGCGCAAGATCATCATGATTAATCAAC 840
DB 980 GACAACTGGGAGCTTCCCATCCCTGCTGTGCGCAAGATCATCATGATTAATCAAC 1039

QY 841 CCCATGTACCCCAAGAACATGACATCGCCCTCATGAAGTGTGAGTGTGAGTGTGAGTGTG 900

RESULT 3
US-10-295-027-133
Sequence 133, Application US/10295027
Publication No. US2003023350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevizi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
Remaining Prior Application data removed - See File Wrapper or PALM.


```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-778

Query Match          99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3  GGATCTGACAGTGATCAACTCTGAAACAGCTGTGATGTAACCCCTGGCGAAACCCCG 62
DB 9  GGATCTGACAGTGATCAACTCTGAAACAGCTGTGATGTAACCCCTGGCGAAACCCCG 68
QY 63  TATCCCATGAGAGACTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
DB 69  TATCCCATGAGAGACTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 128
QY 123  GCGAGATATCATCTATTGTGTCTCTCATCAAGGTGATCTCGATAAATACTACTTCT 182
DB 129  GCGAGATATCATCTATTGTGTGTCTCTCATCAAGGTGATCTCGATAAATACTACTTCT 188
QY 183  CTGCGGGAGCTCTCCACTTCATCCCGAAGACAGCTGTGTACCGAGAGCTGAGCTG 242
DB 189  CTGCGGGAGCTCTCCACTTCATCCCGAAGACAGCTGTGTACCGAGAGCTGAGCTG 248
QY 243  TCCCTTGGGGAGGACGAGGAGCACTGTCTCAAGAGCTTCCCGAAGGGCTCTCAGTGGC 302
DB 249  TCCCTTGGGGAGGACGAGGAGCACTGTCTCAAGAGCTTCCCGAAGGGCTCTCAGTGGC 308
QY 303  AGTCGCGCTCTTCAAGAGACCGATCCACACTCGAGGTGTGTGACATCGGCCACAGGAACTG 362
DB 309  AGTCGCGCTCTTCAAGAGACCGATCCACACTCGAGGTGTGTGACATCGGCCACAGGAACTG 368
QY 363  GTTCTCTGCTGTTTTCGCAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 422
DB 369  GTTCTCTGCTGTTTTCGCAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 428
QY 423  GGGCTACAGCAGCAAAACCCATTTTCAGAGCTGTGTGAGATTTGGCCACAGCAGGATCTGA 482
DB 429  GGGCTACAGCAGCAAAACCCATTTTCAGAGCTGTGTGAGATTTGGCCACAGCAGGATCTGA 488
QY 483  TGTGTGTAATACAGAAACACGACGAGCTTCGATCGGAACTCAGTGGCCCTG 542
DB 489  TGTGTGTAATACAGAAACACGACGAGCTTCGATCGGAACTCAGTGGCCCTG 548
QY 543  TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAAAGCCTGAAGACCC 602
DB 549  TCTCTCAGGCTCCCTGGTCTCCCTGCACTGTCTTGGCTGTGGGAAAGCCTGAAGACCC 608
QY 603  CGGTGTGGTGGGAGGAGGCTCTGTGATTTTGGCTTGGCAGGTGAGATCA 662
DB 609  CGGTGTGGTGGGAGGAGGCTCTGTGATTTTGGCTTGGCAGGTGAGATCA 668
QY 663  GTACCAAAACAGCAGCTCTGTGAGGAGGATCTCTGACCCCACTGGGTCTCTCAGGC 722
DB 669  GTACCAAAACAGCAGCTCTGTGAGGAGGATCTCTGACCCCACTGGGTCTCTCAGGC 728
QY 723  AGCCACTGCTTTCAGAAACATACGATGTGTTCAACTGGAAGGTGGGAGGCTCA 782
DB 729  AGCCACTGCTTTCAGAAACATACGATGTGTTCAACTGGAAGGTGGGAGGCTCA 788
QY 783  CAAACTGGGAGCTTCCCATCCCTGGCTGGCCAGAGATCATCATCAATTAATCAACC 842
DB 789  CAAACTGGGAGCTTCCCATCCCTGGCTGGCCAGAGATCATCATCAATTAATCAACC 848
QY 843  CATGTACCCCAAGACATGATCGCCCTCATGAAGCTCAGTTCCTCACTCTTCTC 902
DB 849  CATGTACCCCAAGACATGATCGCCCTCATGAAGCTCAGTTCCTCACTCTTCTC 908
QY 903  AGGCAAGTTCAGCCCATCTGTCTGCTTCTTGTGATGAGGAGTCACTCAGCACCC 962
DB 909  AGGCAAGTTCAGCCCATCTGTCTGCTTCTTGTGATGAGGAGTCACTCAGCACCC 968
QY 963  ACTCTGGATTCATTGGATGGGCTTTTACAGACAGATGAGGAGGAAAGATGTCTCACATCT 1022

```

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 3 GATCTCTGACATGATCACTCTGACAGACCTCGATGTCACACCCCTGCGCAACCCCG 62
DB 9 GATCTCTGACATGATCACTCTGACAGACCTCGATGTCACACCCCTGCGCAACCCCG 68
QY 63 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCAGCTACTGAGCCT 122
DB 69 TATCCCATGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCAGCTACTGAGCCT 128
QY 123 GGGAGTATCATCTGTTGTTCTCTCATCAAGTGATTCGGAATAATCTACTTCTT 182
DB 129 GGGAGTATCATCTGTTGTTCTCTCATCAAGTGATTCGGAATAATCTACTTCTT 188
QY 183 CTGCGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAATG 242
DB 189 CTGCGGGAGCCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGGAATG 248
QY 243 TCCCTTGGGGAGGACGAGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 302
DB 249 TCCCTTGGGGAGGACGAGAGCAGCTGTGTCAAGAGCTTCCCGAAGGCGCTGAGTGGC 308
QY 303 AGTCGCGCTCTCCAGGACCGATCCACACTGAGCTGTGTGAGCTCGGCCACAGGAACTG 362
DB 309 AGTCGCGCTCTCCAGGACCGATCCACACTGAGCTGTGTGAGCTCGGCCACAGGAACTG 368
QY 363 GTTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 422
DB 369 GTTCTCTGCTGTTTCGACAACTTCACAGAACTCTCGCTGAGACAGCTGTGAGCAGAT 428
QY 423 GGGCTACAGAGAAACCCACTTTGAGAGCTGTGTGAGATGTCGCCACAGGAGTCTGA 482
DB 429 GGGCTACAGAGAAACCCACTTTGAGAGCTGTGTGAGATGTCGCCACAGGAGTCTGA 488
QY 483 TGTGTTGAAATCACAGAAACAGCAGGAGCTTGCATGCGGAATCAAGTGGGCGCTG 542
DB 489 TGTGTTGAAATCACAGAAACAGCAGGAGCTTGCATGCGGAATCAAGTGGGCGCTG 548
QY 543 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGTGCTGTGGGAGAGCTGGAAGCCTC 602
DB 549 TCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTGTGCTGTGGGAGAGCTGGAAGCCTC 608
QY 603 CCGTGTGTTGGTGGGAGAGGCTCTGTGATTTCTGGCTTGGCAGGTGAGCATCCA 662
DB 609 CCGTGTGTTGGTGGGAGAGGCTCTGTGATTTCTGGCTTGGCAGGTGAGCATCCA 668
QY 663 GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTCTGGAACCCCTCTGAGTCTCAGCGC 722
DB 669 GTACGACAAACAGCAGCTGTGTGGAGGAGCATCTCTGGAACCCCTCTGAGTCTCAGCGC 728
QY 723 AGCCCACTGTTCCAGAAACATACCGATGTTTCACTGGAAGTGGCGGAGGCTCAGA 782
DB 729 AGCCCACTGTTCCAGAAACATACCGATGTTTCACTGGAAGTGGCGGAGGCTCAGA 788
QY 783 CAAACTGGGAGCTTCCCATCCCTGCTGTGCTGAGATCATCATGGAATTCAGCC 842
DB 789 CAAACTGGGAGCTTCCCATCCCTGCTGTGCTGAGATCATCATGGAATTCAGCC 848
QY 843 CATGTACCCCAAGAACATGACATGCTGCTCATGAGCTGAGTCCCACTCACTTCTC 902
DB 849 CATGTACCCCAAGAACATGACATGCTGCTCATGAGCTGAGTCCCACTCACTTCTC 908
QY 903 AGGCACAGTCAGGCCCATCTGTCTGCTCTTCTGATGAGGAGCTCACTCCAGCCACCC 962
DB 909 AGGCACAGTCAGGCCCATCTGTCTGCTCTTCTGATGAGGAGCTCACTCCAGCCACCC 968
QY 963 ACTCTGATCATTTGATGAGGCTTTTACGAGAGGAGATGAGGAGATGTTCTGACATCT 1022
DB 969 ACTCTGATCATTTGATGAGGCTTTTACGAGAGGAGATGAGGAGATGTTCTGACATCT 1028
QY 1023 GCTGACAGGCTCAGTCCAGTCTATTACAGACACACCGTCAATGCAGACGATGCTGACCA 1082
```

```
DB 1029 GCTGCAAGGCTCAGTCCAGTCAATGCAGACACACCGTCAATGCAGACGATGCTGACCA 1088
QY 1083 GGGGGAAGTCAACGAGAAGATGATGTGTCAGGACATCCCGAAGGGGTGTGACACCTG 1142
DB 1089 GGGGGAAGTCAACGAGAAGATGATGTGTCAGGACATCCCGAAGGGGTGTGACACCTG 1148
QY 1143 CAGGGTGACAGTGTGCGCCCTGATGTACCAATCTGACCACTGTCATGTGTTGGGCT 1202
DB 1149 CAGGGTGACAGTGTGCGCCCTGATGTACCAATCTGACCACTGTCATGTGTTGGGCT 1208
QY 1203 CGTTAGCTGGGCTGATGTGCTGCGGGGCGCCGAGCACCCAGGAGTATACACCAAGTCTC 1262
DB 1209 CGTTAGCTGGGCTGATGTGCTGCGGGGCGCCGAGCACCCAGGAGTATACACCAAGTCTC 1268
QY 1263 AGCTATCTCACTGATCTACCAATGTCTGGAAGGCTGAGCTG 1305
DB 1269 AGCTATCTCACTGATCTACCAATGTCTGGAAGGCTGAGCTG 1311
```

RESULT 6

US-10-295-027-830
; Sequence 830, Application US/10295027
; Publication No. US20030232350A1

GENERAL INFORMATION:

APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 830
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-295-027-830

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCCTGACAGTGATACCTCTGACAGCCTCGATGTCAAACCCCTCGGCAACCCCG 62
Db 9 GGATCCTGACAGTGATACCTCTGACAGCCTCGATGTCAAACCCCTCGGCAACCCCG 68
QY 63 TATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 69 TATCCCATGGAGACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 128
QY 123 GCGAGATATCATATGTTGGTCTCTCATCAAGGTGATCTGGATAAATACTCTCT 182
Db 129 GCGAGATATCATATGTTGGTCTCTCATCAAGGTGATCTGGATAAATACTCTCTCT 188
QY 183 CTGCGGGAGCCTCTCCACTTCATCCGAGGAGAGAGCTGTGTGACGGAGAGCTGCACTG 242
Db 189 CTGCGGGAGCCTCTCCACTTCATCCGAGGAGAGAGCTGTGTGACGGAGAGCTGCACTG 248
QY 243 TCCCTTGGGGAGGAG 302
Db 249 TCCCTTGGGGAGGAG 308
QY 303 AGTCCGCTCTCCAGGAGCCGATCCACACTGCTGAGCTGCTGAGCTCGGCGACAGGAACTG 362
Db 309 AGTCCGCTCTCCAGGAGCCGATCCACACTGCTGAGCTGCTGAGCTCGGCGACAGGAACTG 368
QY 363 GTTCTCTGCTCTTTCGAACTTCACAGAACTCTCTGCTGAGAGAGCTGTGAGGAGAGAG 422
Db 369 GTTCTCTGCTCTTTCGAACTTCACAGAACTCTCTGCTGAGAGAGCTGTGAGGAGAGAG 428
QY 423 GGGCTACAGCAGCAACCCACTTTCAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
Db 429 GGGCTACAGCAGCAACCCACTTTCAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 483 TGTGTGTGAATTCACAGAAACAGCAGGAGCTTCGCTGAGCTCGGAACTCAAGTGGGCCCTG 542
Db 489 TGTGTGTGAATTCACAGAAACAGCAGGAGCTTCGCTGAGCTCGGAACTCAAGTGGGCCCTG 548
QY 543 TCTCTCAGCTCCTGCTCTCCCTGCACTGTCTTCTGCTGTGGGAGAGAGCTGTGAGAGAGAG 602
Db 549 TCTCTCAGCTCCTGCTCTCCCTGCACTGTCTTCTGCTGTGGGAGAGAGCTGTGAGAGAGAG 608
QY 603 CCGTGTGTGGTGGGAGAGAGAGCTCTGTGATTTCTGGCTTCGAGAGTTCAGCATCA 662
Db 609 CCGTGTGTGGTGGGAGAGAGAGCTCTGTGATTTCTGGCTTCGAGAGTTCAGCATCA 668
QY 663 GTACGACAAACAGCAGCTCTGTGGAGGAGAGATCTGAGACCCCACTGGGTCTCA 722
Db 669 GTACGACAAACAGCAGCTCTGTGGAGGAGAGATCTGAGACCCCACTGGGTCTCA 728
QY 723 AGCCCACTCTTCAGGAAACATACCGATGTGTTCAGTGGAGGTGCGGAGAGGCTCAGA 782
Db 729 AGCCCACTCTTCAGGAAACATACCGATGTGTTCAGTGGAGGTGCGGAGAGGCTCAGA 788
QY 783 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACACCC 842
Db 789 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATATTGAATTCACACCC 848
QY 843 CATGTACCCCAAGACATGATGCCCTCATGAGCTGAGTTCACCACTCATCTTCTC 902
Db 849 CATGTACCCCAAGACATGATGCCCTCATGAGCTGAGTTCACCACTCATCTTCTC 908
QY 903 AGGCAAGTCAGGAGGAGCTGTCTGCTGCTCTTTGATGAGGAGCTCACTCCAGGAGAGAG 962
Db 909 AGGCAAGTCAGGAGGAGCTGTCTGCTGCTCTTTGATGAGGAGCTCACTCCAGGAGAGAG 968
QY 963 ACTCTGGATCATGATGGGCTTTTACGAAGCAGAGATGAGGAGAGAGATGTCTGACATCT 1022
Db 969 ACTCTGGATCATGATGGGCTTTTACGAAGCAGAGATGAGGAGAGAGATGTCTGACATCT 1028
QY 1023 GCTGAGGAGCTGAGTCCAGGTTCATGACAGCAGAGCTGCAATGACAGAGATGCTGACCA 1082
Db 1029 GCTGAGGAGCTGAGTCCAGGTTCATGACAGCAGAGCTGCAATGACAGAGATGCTGACCA 1088
QY 1083 GGGGGAAGTCACCGAGAGAGATGATGTGTGAGGAGCTCCGGAAGGGGGTGTGGACACCTG 1142

Db 1089 GGGGGAAGTCACCGAGAGAGATGATGTGTGAGGAGCTCCGGAAGGGGTGTGACACCTG 1148
QY 1143 CCAGGGTGACAGTGGTGGGCCCTGTATGTATACCAATCTGACAGTGGCATGTGTGGGCAT 1202
Db 1149 CCAGGGTGACAGTGGTGGGCCCTGTATGTATACCAATCTGACAGTGGCATGTGTGGGCAT 1208
QY 1203 CGTTAGCTGGGCTATGGCTGGGGGCCCGAGCAGCAGGAGATATACACCAAGGTCTC 1262
Db 1209 CGTTAGCTGGGCTATGGCTGGGGGCCCGAGCAGCAGGAGATATACACCAAGGTCTC 1268
QY 1263 AGCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1305
Db 1269 AGCTATCTCAACTGGATCTACAAATGTCTGGAAGGCTGAGCTG 1311

RESULT 7

US-10-295-027-979
; Sequence 979, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 979
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-979

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATCCTGACAGTGATCACTCTGACAGCCTCGATGTCAAACCCCTCGGCAACCCCG 62
Db 9 GGATCCTGACAGTGATCACTCTGACAGCCTCGATGTCAAACCCCTCGGCAACCCCG 68

QY 63 TATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATATAGCACTACTGAGCCT 122
DB 69 TATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATATAGCACTACTGAGCCT 128
QY 123 GGGAGATATCATATGTGGTGTCTCTCATCAAGGTGATCTGGATTAATACTACTTCT 182
DB 129 GGGAGATATCATATGTGGTGTCTCTCATCAAGGTGATCTGGATTAATACTACTTCT 188
QY 183 CTGGGGAGACCTTCCACTTCATCCGAGGAGAGCTGTGTGACGAGAGCTGGACTG 242
DB 189 CTGGGGAGACCTTCCACTTCATCCGAGGAGAGCTGTGTGACGAGAGCTGGACTG 248
QY 243 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 302
DB 249 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 308
QY 303 AGTCCGCTCTCCAGGACCGATCCACACTGAGGTGCTGGACTCGGCACAGGAACTG 362
DB 309 AGTCCGCTCTCCAGGACCGATCCACACTGAGGTGCTGGACTCGGCACAGGAACTG 368
QY 363 GTTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAT 422
DB 369 GTTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAT 428
QY 423 GGGCTACAGAGAAACCCACTTTCAGAGCTGTGAGATGGCCACAGCAGGATCTGA 482
DB 429 GGGCTACAGAGAAACCCACTTTCAGAGCTGTGAGATGGCCACAGCAGGATCTGA 488
QY 483 TGTGTGTAATCACAGAAAACAGCAGGAGCTTCGATGCGGAATCAAGTGGGCCCTG 542
DB 489 TGTGTGTAATCACAGAAAACAGCAGGAGCTTCGATGCGGAATCAAGTGGGCCCTG 548
QY 543 TCTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
DB 549 TCTCTCAGGCTCCCTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 608
QY 603 CCGTGTGGTGGGGAGGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662
DB 609 CCGTGTGGTGGGGAGGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 668
QY 663 GTACGACAAACAGCAGCTGTGTGGAGGAGAGCTTCCTGGAACCCCTCTGCTGCTGCTG 722
DB 669 GTACGACAAACAGCAGCTGTGTGGAGGAGAGCTTCCTGGAACCCCTCTGCTGCTGCTG 728
QY 723 AGCCGACTGTTTACGAAATACCGATGTTTCACTGGAAGGTGCGGAGGCTCAGA 782
DB 729 AGCCGACTGTTTACGAAATACCGATGTTTCACTGGAAGGTGCGGAGGCTCAGA 788
QY 783 CAAACTGGGAGCTTCCCATCCCTGCTGTGGCCCAAGATCATCATTTGAATTCACACC 842
DB 789 CAAACTGGGAGCTTCCCATCCCTGCTGTGGCCCAAGATCATCATTTGAATTCACACC 848
QY 843 CATGTACCCCAAGACAATGACATCCCTCTCATGAGCTGAGTTCCTCACTTCTC 902
DB 849 CATGTACCCCAAGACAATGACATCCCTCTCATGAGCTGAGTTCCTCACTTCTC 908
QY 903 AGGCAAGTTCAGGCCCATCTGTCTGCTGCTTCTTGTATGAGAGCTCACTTCAGCCACCC 962
DB 909 AGGCAAGTTCAGGCCCATCTGTCTGCTGCTTCTTGTATGAGAGCTCACTTCAGCCACCC 968
QY 963 ACTCTGGATCATTTGATGGGCTTTACGAAGCAGATGAGGAGAGATGTCTGACATCT 1022
DB 969 ACTCTGGATCATTTGATGGGCTTTACGAAGCAGATGAGGAGAGATGTCTGACATCT 1028
QY 1023 GCTGAGGAGTTCAGTCCAGTTCATTCAGACAGCACCGGTGCAATGAGAGAGTTCGATCCA 1082
DB 1029 GCTGAGGAGTTCAGTCCAGTTCATTCAGACAGCACCGGTGCAATGAGAGAGTTCGATCCA 1088
QY 1083 GGGGGAAGTTCACGAGAAATGATGTGTGAGGATCCCGAAGGGGTGTGACACCTG 1142
DB 1089 GGGGGAAGTTCACGAGAAATGATGTGTGAGGATCCCGAAGGGGTGTGACACCTG 1148
QY 1143 CAGGGGTGACGTGGTGGGGCCCTGTATGATACCAATCTGACAGTGGCATGTGGTGGCAT 1202

DB 1149 CCAGGGTGACAGTGGTGGGCCCTGTATGTACCAATCTGCAGTGGCATGTGGTGGCAT 1208
QY 1203 CGTTAGCTGGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACCAAGGTCTC 1262
DB 1209 CGTTAGCTGGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACCAAGGTCTC 1268
QY 1263 AGCTATCTCAACTGGATCTACAAATCTCTGGAAGGTGAGCTG 1305
DB 1269 AGCTATCTCAACTGGATCTACAAATCTCTGGAAGGTGAGCTG 1311

RESULT: 8
US-10-173-999-88
; Sequence 88, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-173-999-88

Query Match 99.8%; Score 1303; DB 16; Length 1314;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCAACCCCTCGCAAAACCCCG 62
DB 9 GGATCTGACAGTATCAACCTCTGACAGCCTCGATGTCAACCCCTCGCAAAACCCCG 68
QY 63 TATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATATAGCACTACTGAGCCT 122
DB 69 TATCCCATGGAGACCTTCAGAAAGGTGGGGATCCCATCATATAGCACTACTGAGCCT 128
QY 123 GGGAGTATCATATGTGGTGTCTCTCATCAAGGTGATCTGGATTAATACTACTTCT 182
DB 129 GGGAGTATCATATGTGGTGTCTCTCATCAAGGTGATCTGGATTAATACTACTTCT 188
QY 183 CTGGGGAGGACCTTCACATTTTCATCCGAGGAGAGCTGTGTGACGAGAGCTGGACTG 242
DB 189 CTGGGGAGGACCTTCACATTTTCATCCGAGGAGAGCTGTGTGACGAGAGCTGGACTG 248
QY 243 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 302
DB 249 TCCCTTGGGGAGGACGAGGACACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 308
QY 303 AGTCCGCTCTCCAGGACCGATCCACACTGAGGTGCTGGACTCGGCACAGGAACTG 362
DB 309 AGTCCGCTCTCCAGGACCGATCCACACTGAGGTGCTGGACTCGGCACAGGAACTG 368
QY 363 GTTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAT 422
DB 369 GTTCTCTGCTGTTCGACAACTTCACAGAGCTCTCGCTGAGACAGCTGTAGGACAT 428

QY 423 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGAGATTGGCCAGACAGAGATCTCGA 482
DB 429 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGAGATTGGCCAGACAGAGATCTCGA 488
QY 483 TGTGTGTGAATTCACAGAAACAGCAGAGCTTCGATGGGAACTCAAGTGGGCGCTG 542
DB 489 TGTGTGTGAATTCACAGAAACAGCAGAGCTTCGATGGGAACTCAAGTGGGCGCTG 548
QY 543 TCTCTCAGGCTCCCTGGTCTCCCTCAGCTGTCTGCTGTGGGAAAGAGCCTGAAGACCC 602
DB 549 TCTCTCAGGCTCCCTGGTCTCCCTCAGCTGTCTGCTGTGGGAAAGAGCCTGAAGACCC 608
QY 603 CGGTGTGGTGGGGAGGAGGCTCTGTGAGTCTTGGCTTGGCAGGTCAGATCCA 662
DB 609 CGGTGTGGTGGGGAGGAGGCTCTGTGAGTCTTGGCTTGGCAGGTCAGATCCA 668
QY 663 GTACGACAAACAGCAGCTGTGGAGGAGAGATCTCGACCCCACTGGCTCCTCAGGCG 722
DB 669 GTACGACAAACAGCAGCTGTGGAGGAGAGATCTCGACCCCACTGGCTCCTCAGGCG 728
QY 723 AGCCCACTGCTCAGAAACATACCGATGTGTAACTGGAGGTGGGGAGGCTCAGA 782
DB 729 AGCCCACTGCTCAGAAACATACCGATGTGTAACTGGAGGTGGGGAGGCTCAGA 788
QY 783 CAAACTGGGAGCTTCCCACTCCCTGGCTGTGGCCAGAGATCATCATTTGAATCAACCC 842
DB 789 CAAACTGGGAGCTTCCCACTCCCTGGCTGTGGCCAGAGATCATCATTTGAATCAACCC 848
QY 843 CATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTTCACACTTTCTC 902
DB 849 CATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGCAGTTCACACTTTCTC 908
QY 903 AGGCACAGTCAGGCCCATCTGTCTGCCCTTTTGTGATGAGAGCTCACTCCAGCCACCC 962
DB 909 AGGCACAGTCAGGCCCATCTGTCTGCCCTTTTGTGATGAGAGCTCACTCCAGCCACCC 968
QY 963 ACTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAATGGAGGAGAGATGTCTGACATACT 1022
DB 969 ACTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAATGGAGGAGAGATGTCTGACATACT 1028
QY 1023 GCTGAGGCTCAGTCCAGGTCATTGACAGACACAGTGGTGAATGAGAGATGGGTACCA 1082
DB 1029 GCTGAGGCTCAGTCCAGGTCATTGACAGACACAGTGGTGAATGAGAGATGGGTACCA 1088
QY 1083 GGGGAAGTCAACGAGAGATGATGTGAGGATCCCGAAGGGGTGTGGACACTG 1142
DB 1089 GGGGAAGTCAACGAGAGATGATGTGAGGATCCCGAAGGGGTGTGGACACTG 1148
QY 1143 CCAGGTGACAGTGTGGGCCCTTGATGTACCAATCTGACAGTGGCATGTGGGGCAT 1202
DB 1149 CCAGGTGACAGTGTGGGCCCTTGATGTACCAATCTGACAGTGGCATGTGGGGCAT 1208
QY 1203 CGTTAGCTGGGCTATGCTGGGGCCCGAGCAGCCAGGATATACACAGGTC 1262
DB 1209 CGTTAGCTGGGCTATGCTGGGGCCCGAGCAGCCAGGATATACACAGGTC 1268
QY 1263 AGCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1305
DB 1269 AGCTATCTCAACTGGATCTCAATGTCTGGAAGGCTGAGCTG 1311

RESULT 9

US-09-851-588-7
; Sequence 7, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS
; FILE REFERENCE: A-68829-1/DJB/JUD/AMS

; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,252
; PRIOR FILING DATE: 2000-08-17
; CURRENT APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 2081
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (215)..(1528)
; OTHER INFORMATION:
US-09-851-588-7

Query Match 99.8%; Score 1303; DB 9; Length 2081;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATCTCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAAAACCCCG 62
DB 223 GGATCTCTGACAGTGATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGCGCAAAACCCCG 282
QY 63 TATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGCCT 122
DB 283 TATCCCCATGGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTAGCCT 342
QY 123 GGGAGATATCATATTGTGGTGTCTCATCAAGGTGATTCTGATATAAATACTACTTCT 182
DB 343 GGGAGATATCATATTGTGGTGTCTCATCAAGGTGATTCTGATATAAATACTACTTCT 402
QY 183 CTGGGGAGCCTCTCCATTTTCATCCCGAGGAAGAGCTGTGTGACGAGAGTGGACTG 242
DB 403 CTGGGGAGCCTCTCCATTTTCATCCCGAGGAAGAGCTGTGTGACGAGAGTGGACTG 462
QY 243 TCCTTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTGGC 302
DB 463 TCCTTTGGGGAGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCTTCAGTGGC 532
QY 303 AGTCGCGCTCTCCAAAGGACCGATCCACACTGCAAGTGTGTGACTCGGCCACAGGAACTG 362
DB 523 AGTCGCGCTCTCCAAAGGACCGATCCACACTGCAAGTGTGTGACTCGGCCACAGGAACTG 582
QY 363 GTTCTCTGCTGTTTTCGACAACTTTCAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 422
DB 583 GTTCTCTGCTGTTTTCGACAACTTTCAGAGCTCTCGCTGAGACAGCTGTAGGAGAT 642
QY 423 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 482
DB 643 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCAGACAGGATCTGGA 702
QY 483 TGTGTGTGAATTCACAGAAACAGCAGAGCTTGCATGCGGAACTCAAGTGGGCGCTG 542
DB 703 TGTGTGTGAATTCACAGAAACAGCAGAGCTTGCATGCGGAACTCAAGTGGGCGCTG 762
QY 543 TCTCTCAGGCTCCCTGGTCTCCCTGACATGTCTGCTGAGAGAGCTTGAAGACCC 602
DB 763 TCTCTCAGGCTCCCTGGTCTCCCTGACATGTCTGCTGAGAGAGCTTGAAGACCC 822
QY 603 CGGTGTGGTGGGGAGGAGGCTCTCTGTGATTTTGGCTTGGCAGGTGAGCATCCA 662
DB 823 CGGTGTGGTGGGGAGGAGGCTCTCTGTGATTTTGGCTTGGCAGGTGAGCATCCA 882
QY 663 GTACGACAAACAGCAGCTCTGTGGAGGAGATCTTGGACCCCACTGGGTCTCAGGCG 722
DB 883 GTACGACAAACAGCAGCTCTGTGGAGGAGATCTTGGACCCCACTGGGTCTCAGGCG 942
QY 723 AGCCCACTGCTCAGAAACATACCGATGTGTAACTGGAAGGTGCGGGAGGCTCAGA 782
DB 943 AGCCCACTGCTCAGAAACATACCGATGTGTAACTGGAAGGTGCGGGAGGCTCAGA 1002

783 CAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATGATTAATCAACCC 842
1003 CAACTGGGAGCTTCCCATCCCTGGCTGTGGCCCAAGATCATCATGATTAATCAACCC 1062
843 CATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGGAGTCCCACTTCTC 902
1063 CATGTACCCCAAGACAATGACATCGCCCTCATGAAGTGGAGTCCCACTTCTC 1122
903 AGGCAGTCCAGCCCATCTGCTGCCCTTCTTTCATGAGGAGCTCACTCCAGCCACCC 962
1123 AGGCAGTCCAGCCCATCTGCTGCCCTTCTTTCATGAGGAGCTCACTCCAGCCACCC 1182
963 ACTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1022
1183 ACTCTGGATCATTTGATGGGGCTTTACGAAGCAGAAATGAGGGAAGATGTCTGACATACT 1242
1023 GGTGAGGCGTCACTCCAGTCAATTCAGACGACACAGGTCATGAGGAGTGGTACCA 1082
1243 GGTGAGGCGTCACTCCAGTCAATTCAGACGACACAGGTCATGAGGAGTGGTACCA 1302
1083 GGGGAAGTCACCGAAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACCTG 1142
1303 GGGGAAGTCACCGAAGATGATGTGTCAGGCAATCCCGAAGGGGTGTGGACACCTG 1362
1143 CCAGGTCAGTGTGGGCGCTGATGTACCAATCTGACAGTGGCATGTGGGGCAT 1202
1363 CCAGGTCAGTGTGGGCGCTGATGTACCAATCTGACAGTGGCATGTGGGGCAT 1422
1203 CGTTAGCTGGGCTATGGCTGGGGCGGCGGACACCCAGGAGTATACCAAGGTCTC 1262
1423 CGTTAGCTGGGCTATGGCTGGGGCGGCGGACACCCAGGAGTATACCAAGGTCTC 1482
1263 AGCTATCTCAACTGATCTACAAATGTCTGGAAGCTGAGCTG 1305
1483 AGCTATCTCAACTGATCTACAAATGTCTGGAAGCTGAGCTG 1525

RESULT 10
US-10-097-340-317
; Sequence 317, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISBY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. EAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US/10/097,340
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732

;; PRIOR FILING DATE: 2001-08-10
;; PRIOR APPLICATION NUMBER: 60/325,102
;; PRIOR FILING DATE: 2001-09-26
;; PRIOR APPLICATION NUMBER: 60/323,580
;; PRIOR FILING DATE: 2001-09-19
;; NUMBER OF SEQ ID NOS: 363
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 317
;; LENGTH: 2307
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(2307)
;; OTHER INFORMATION: n = A,T,C or G
US-10-097-340-317

Query Match 99.8%; Score 1303; DB 15; Length 2307;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GGATCTCGACAGTGAATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGGGCAACCCCG 62
Db 283 GGATCTCGACAGTGAATCAACCTCTGAACAGCCTCGATGTCAAACCCCTGGGCAACCCCG 342
Qy 63 TATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 343 TATCCCCATGAGACCTTCAGAAAGGTGGGATCCCATCATCATAGCACTACTGAGCCT 402
Qy 123 GGGAGTATCATCATTTGTGTTCTCTCATCAAGTGTGTTGGTAAATACTACTTCT 182
Db 403 GGGAGTATCATCATTTGTGTTCTCTCATCAAGTGTGTTGGTAAATACTACTTCT 462
Qy 183 CTGGGGCAGCCTCTCCATTCATCCGAGGAGCAGCTGTGTGCGGAGAGCTGACTG 242
Db 463 CTGGGGCAGCCTCTCCATTCATCCGAGGAGCAGCTGTGTGCGGAGAGCTGACTG 522
Qy 243 TCCCTTGGGGGAGGACGAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGGC 302
Db 523 TCCCTTGGGGGAGGACGAGGAGCAGTGTGTCAAGAGCTTCCCGAAGGGCTGCACTGGC 582
Qy 303 AGTCCGCTCTCAAGAGCAGATTCACATGAGGTCGTGGATCGGCTGGCCACAGGAACTG 362
Db 583 AGTCCGCTCTCAAGAGCAGATTCACATGAGGTCGTGGATCGGCTGGCCACAGGAACTG 642
Qy 363 GTTCTGCTGCTGTTTCGACAACTTCAAGAGCTCTCGCTGACACAGCCTGTAGGAGAT 422
Db 643 GTTCTGCTGCTGTTTCGACAACTTCAAGAGCTCTCGCTGACACAGCCTGTAGGAGAT 702
Qy 423 GGGCTACAGCAGCAAAACCCCTTTTCAGAGCTGTGGAGATTTGGCCACAGACAGGATCTGA 482
Db 703 GGGCTACAGCAGCAAAACCCCTTTTCAGAGCTGTGGAGATTTGGCCACAGACAGGATCTGA 762
Qy 483 TGTGTTGAAATCACAGAAACAGCAGGAGCTTCCCATGCGGAACTCAAGTGGGCTG 542
Db 763 TGTGTTGAAATCACAGAAACAGCAGGAGCTTCCCATGCGGAACTCAAGTGGGCTG 822
Qy 543 TCTCTCAGGCTCCCTGCTTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
Db 823 TCTCTCAGGCTCCCTGCTTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 882
Qy 603 CCGTGTGGTGGGAGGAGGAGGCTTGTGATTTCTGGCCTTGGCAGTTCAGCATCCA 662
Db 883 CCGTGTGGTGGGAGGAGGAGGCTTGTGATTTCTGGCCTTGGCAGTTCAGCATCCA 942
Qy 663 GTACCAAAACAGCAGCAGTCTGTGGAGGAGCAGTCTGGAGAGGAGGAGGAGGAGGAGGAGGAG 722
Db 943 GTACCAAAACAGCAGCAGTCTGTGGAGGAGCAGTCTGGAGAGGAGGAGGAGGAGGAGGAGGAG 1002
Qy 723 AGCCCACTGCTTACGAAACATACCCGATGTTCACACTGGAAGTGGGGCAGGCTCAGA 782
Db 1003 AGCCCACTGCTTACGAAACATACCCGATGTTCACACTGGAAGTGGGGCAGGCTCAGA 1062

QY 783 CAAATGGGCGAGTCCCTCCCTGGCTGGTGGCCAGATCATCATCATGTAATCAACCC 842
Db 1063 CAAATGGGCGAGTCCCTCCCTGGCTGGTGGCCAGATCATCATCATGTAATCAACCC 1122
QY 843 CATGTACCCCAAGACAATGACATCGCCCTCATGAGCTGCAAGTTCACCACTCCTCTC 902
Db 1123 CATGTACCCCAAGACAATGACATCGCCCTCATGAGCTGCAAGTTCACCACTCCTCTC 1182
QY 903 AGGCAGTTCAGGCGGATGCTGCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCC 962
Db 1183 AGGCAGTTCAGGCGGATGCTGCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCC 1242
QY 963 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGAGGGAAGATGCTGCATACT 1022
Db 1243 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGAGGGAAGATGCTGCATACT 1302
QY 1023 GCTGAGGCGTTCAGTCCAGGTCAATGACAGACACAGTGGTCAATGACAGATGCGTACCA 1082
Db 1303 GCTGAGGCGTTCAGTCCAGGTCAATGACAGACACAGTGGTCAATGACAGATGCGTACCA 1362
QY 1083 GGGGAAGTCAACGAGAGATGATGTGCGAGGATCCCGAAGGGGTGCGACACTG 1142
Db 1363 GGGGAAGTCAACGAGAGATGATGTGCGAGGATCCCGAAGGGGTGCGACACTG 1422
QY 1143 CCAGGTCAGAGTGGGCGGCTCATGTACCAATCTGACCAATCTGACCAATCTGACCAATCTG 1202
Db 1423 CCAGGTCAGAGTGGGCGGCTCATGTACCAATCTGACCAATCTGACCAATCTGACCAATCTG 1482
QY 1203 GCTTAGCTGGGCTATGCTGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1262
Db 1483 GCTTAGCTGGGCTATGCTGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1542
QY 1263 AGCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1305
Db 1543 AGCTATCTCAACTGGATCTACAATGTCTGGAAGGCTGAGCTG 1585

RESULT 11

US-10-171-311-217
; Sequence 217, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumel
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerish, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217
; LENGTH: 2307
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 1693, 1698, 1705, 1708, 1709, 1713, 1717, 1720, 1724, 1728,
; LOCATION: 1733, 1741, 1746, 1748, 1755, 1770, 1774, 1791, 1802, 1821,
; LOCATION: 1838, 1856, 1859, 1864, 1908, 1959, 1997, 2012, 2038, 2143

; OTHER INFORMATION: n = A, T, C or G
US-10-171-311-217

Query Match 99.8%; Score 1303; DB 15; Length 2307;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATCTCGACAGTGTCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 62
Db 283 GGATCTCGACAGTGTCAACCTCTGAACAGCTCGATGTCAAAACCCCTGCGCAACCCCG 342
QY 63 TATCCCATCGAGACCTTTAGAAAGGTGGGATCCCATCATCATAGCACTACTAGAGCT 122
Db 343 TATCCCATCGAGACCTTTAGAAAGGTGGGATCCCATCATCATAGCACTACTAGAGCT 402
QY 123 GGCAGTATCATCATTTGTGTTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 182
Db 403 GGCAGTATCATCATTTGTGTTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 462
QY 183 CTGGGGCAGCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGACGAGAGTGGACTG 242
Db 463 CTGGGGCAGCTCTCCACTTTCATCCGAGGAGCAGCTGTGTGACGAGAGTGGACTG 522
QY 243 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 302
Db 523 TCCCTTGGGGAGGACGAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCTTGCAGTGGC 582
QY 303 AGTCGGCTCTCCAAGACCGATCCACATGCAAGTGTGGACTCGGCCACAGGGAACTG 362
Db 583 AGTCGGCTCTCCAAGACCGATCCACATGCAAGTGTGGACTCGGCCACAGGGAACTG 642
QY 363 GTTCTCTGCTGTTTTCGACAACTTTCAGAAAGCTCTCGCTGAGACAGCCCTGTAGGAGAT 422
Db 643 GTTCTCTGCTGTTTTCGACAACTTTCAGAAAGCTCTCGCTGAGACAGCCCTGTAGGAGAT 702
QY 423 GGGCTACAGCAGCAAAACCCACTTTAGAGCTGTGGAGATTTGGCCCAAGACAGATCTGGA 482
Db 703 GGGCTACAGCAGCAAAACCCACTTTAGAGCTGTGGAGATTTGGCCCAAGACAGATCTGGA 762
QY 483 TGTGTTGAAATCACAGAAACAGCAGGAGCTTCCATGCGGAACTCAAGTGGGCGCTG 542
Db 763 TGTGTTGAAATCACAGAAACAGCAGGAGCTTCCATGCGGAACTCAAGTGGGCGCTG 822
QY 543 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAAAGAGCTGAAGACCCC 602
Db 823 TCTCTAGGCTCCCTGCTCTCCCTGCACTGTCTTGGCTGTGGAAAGAGCTGAAGACCCC 882
QY 603 CCGTGTGGTGGGAGGAGGAGGCTCTGTTGATTTCTTGGCTTGGCAGTCAAGTCCCA 662
Db 883 CCGTGTGGTGGGAGGAGGAGGCTCTGTTGATTTCTTGGCTTGGCAGTCAAGTCCCA 942
QY 663 GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGGACCCCATCTGAGTCTCAGGC 722
Db 943 GTACGACAAACAGCAGCTCTGTGGAGGAGCATCTCGGACCCCATCTGAGTCTCAGGC 1002
QY 723 AGCCCACTGTTTACGAAACATACCGATGTTTCAACTGGAAGTGGGCGAGGCTCAGA 782
Db 1003 AGCCCACTGTTTACGAAACATACCGATGTTTCAACTGGAAGTGGGCGAGGCTCAGA 1062
QY 783 CAAACTGGGCGAGTTCCTCCCTGCTGTGGCCAAAGATCATCATATTGAATTCACCC 842
Db 1063 CAAACTGGGCGAGTTCCTCCCTGCTGTGGCCAAAGATCATCATATTGAATTCACCC 1122
QY 843 CATGTACCCCAAGACAATGATCCCTCATGAGCTGCAAGTTCCTCACTACTTCTC 902
Db 1123 CATGTACCCCAAGACAATGATCCCTCATGAGCTGCAAGTTCCTCACTACTTCTC 1182
QY 903 AGGCAGTTCAGGCGGATGCTGCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCC 962
Db 1183 AGGCAGTTCAGGCGGATGCTGCTGCCCTCTTTGATGAGGAGTCACTCCAGCCACCC 1242
QY 963 ACTCTGATCATTTGGATGGGCTTTTACGAAGCAGAAATGAGGGAAGATGCTGACATCT 1022

1243 ACTCTGGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAAGATGTCTGACATACT 1302
1023 GGTGAGGGGTGAGTCCAGTCAATTTGACAGACACGGTGCATTCAGACAGGATGGTACCA 1082
1303 GGTGAGGGGTGAGTCCAGTCAATTTGACAGACACGGTGCATTCAGACAGGATGGTACCA 1362
1083 GGGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACACCTG 1142
1363 GGGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACACCTG 1422
1143 CCAGGTGACAGTGTGGGCTCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGCAT 1202
1423 CCAGGTGACAGTGTGGGCTCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGCAT 1482
1203 CGTTAGCTGGGCTATGGCTGGGGGGCCCGAGACCCCGAGAGATATACCAAGGTCTC 1262
1483 CGTTAGCTGGGCTATGGCTGGGGGGCCCGAGACCCCGAGAGATATACCAAGGTCTC 1542
1263 AGCTATCTCAACTGGATCTACAATGTCTGGAAGCTGAGCTG 1305
1543 AGCTATCTCAACTGGATCTACAATGTCTGGAAGCTGAGCTG 1585

RESULT 12
US-09-776-191-3
; Sequence 3, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Chn
; APPLICANT: Jiumn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2137
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261)...(1574)
; OTHER INFORMATION: DNA sequence encoding a transmembrane serine
; OTHER INFORMATION: protease (MTSP3) protein
US-09-776-191-3

Query Match 99.5%; Score 1298.2; DB 10; Length 2137;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 GATCTGACAGTGTATCACTCTGAACAGCCCTGATGTCAAAACCCCTGCGCAAAACCCG 62
Db 269 GATCTGACAGTGTATCACTCTGAACAGCCCTGATGTCAAAACCCCTGCGCAAAACCCG 328
QY 63 TATCCCCATGGACACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 122
Db 329 TATCCCCATGGACACCTTCAGAAAGTGGGGATCCCATCATCATAGCACTACTGAGCCT 388

QY 123 GGCAGATATCATCATTTGGTGTCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 182
Db 389 GGCAGATATCATCATTTGGTGTCTCTCATCAAGGTGATTTCTGGATAAATACTACTTCT 448
QY 183 CTGGGGGACCTCTCCACTTCATCCCGAGGAGAGAGCTGTGTGACGAGAGCTGAGCTG 242
Db 449 CTGGGGGACCTCTCTCATTTATCCCGAGGAGAGAGCTGTGTGACGAGAGCTGAGCTG 508
QY 243 TCCCTTTGGGGAGGAGAGAGAGCACTGTGTCAAGAGCTTTCCCGAAGGCGCTTCCAGTGGC 302
Db 509 TCCCTTTGGGGAGGAGAGAGAGCACTGTGTCAAGAGCTTTCCCGAAGGCGCTTCCAGTGGC 568
QY 303 AGTCCGCTCTCCAAAGACCGATCCACATCGACAGTGTGTGAGTCCGCGACACAGGAGCTG 362
Db 569 AGTCCGCTCTCCAAAGACCGATCCACATCGACAGTGTGTGAGTCCGCGACACAGGAGCTG 628
QY 363 GTTCTCTGCTGCTTTTCGACAACTTTCAGAGAGCTCTCGCTGAGACAGCTCTAGGACAGAT 422
Db 629 GTTCTCTGCTGCTTTTCGACAACTTTCAGAGAGCTCTCGCTGAGACAGCTCTAGGACAGAT 688
QY 423 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGTGAGATTTGCCCCAGACAGAGATCTGA 482
Db 689 GGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGTGAGATTTGCCCCAGACAGAGATCTGA 748
QY 483 TGTGTGTGAATCACAGAAACAGCCAGGAGCTTCCGATGCGGAATCAAGTGGGCGCTG 542
Db 749 TGTGTGTGAATCACAGAAACAGCCAGGAGCTTCCGATGCGGAATCAAGTGGGCGCTG 808
QY 543 TCTCTCAGGCTCCCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
Db 809 TCTCTCAGGCTCCCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 868
QY 603 CCGTGTGGTGGTGGGAGAGAGCTCTGTGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 662
Db 869 CCGTGTGGTGGTGGGAGAGAGCTCTGTGAGTCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 928
QY 663 GTACGACAAACAGCAGCTCTGTGTGAGGAGAGCTCTGTGAGAGAGAGCTCTGTGAGAGAGCT 722
Db 929 GTACGACATACAGCAGCTCTGTGTGAGGAGAGCTCTGTGAGAGAGAGCTCTGTGAGAGAGCT 988
QY 723 AGCCCACTGCTTCCAGAAACATACCATGTGTCAACTGGAAGTGGGAGAGCTTGGCTTGGCTTGGCT 782
Db 989 AGCCCACTGCTTCCAGAAACATACCATGTGTCAACTGGAAGTGGGAGAGCTTGGCTTGGCTTGGCT 1048
QY 783 CAAACTGGGAGCTTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
Db 1049 CAAACTGGGAGCTTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108
QY 843 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTTCCTGCTGCTGCTGCTGCTGCT 902
Db 1109 CATGTACCCCAAGACAAATGACATCGCCCTCATGAAGCTGAGTTCCTGCTGCTGCTGCTGCTGCT 1168
QY 903 AGCAGAGTCAAGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962
Db 1169 AGCAGAGTCAAGCCCATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
QY 963 ACTCTGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAGAGATGTCTGACATCT 1022
Db 1229 ACTCTGATCATTTGGATGGGCTTTACGAAGCAGAAATGGAGGAGAGATGTCTGACATCT 1288
QY 1023 GCTCAGGCTGAGTCCAGGTCTTGTGACAGCAGACGCTGCAATGACAGAGCTGCGTACCA 1082
Db 1289 GCTCAGGCTGAGTCCAGGTCTTGTGACAGCAGACGCTGCAATGACAGAGCTGCGTACCA 1348
QY 1083 GGGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACACCTG 1142
Db 1349 GGGGGAAGTCAACGAGAGATGATGTGTGAGGATCCCGGAAGGGGTGTGGACACCTG 1408
QY 1143 CCAGGTGACAGTGTGGGCTCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGCAT 1202
Db 1409 CCAGGTGACAGTGTGGGCTCCCTGATGTACCAATCTGACAGTGGCATGTGGTGGCAT 1468
QY 1203 CGTTAGCTGGGCTATGGCTGGGGGGCCCGAGACCCCGAGAGATATACCAAGGTCTC 1262

Db 1469 CGTTAGCTGGGCTATGGCTGGGGGGCCGAGCACCCAGGAGTATACCAAGGTCTC 1528
QY 1263 AGCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG 1305
Db 1529 AGCCTATCTCAACTGGATCTCAATGTCTGGAGGCTGAGCTG 1571

RESULT 13

US-10-156-214A-3
; Sequence 3, Application US/10156214A
; Publication No. US20040001801A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Joseph Edward Semple
; APPLICANT: George P. Vlasuk
; APPLICANT: Scott Jeffrey Kemp
; APPLICANT: Mallareddy Komandla
; APPLICANT: Daniel Vanna Siev
; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use
; FILE REFERENCE: 24745-1611
; CURRENT APPLICATION NUMBER: US/10/156,214A
; CURRENT FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2137
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261)...(1574)
; OTHER INFORMATION: Nucleic acid encoding a transmembrane serine
; OTHER INFORMATION: protease (MSP3) protein
US-10-156-214A-3

Query Match 99.5%; Score 1298.2; DB 16; Length 2137;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1300; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCAACCCCG 62
Db 269 GGATCCTGACAGTATCAACCTCTGAACAGCCTCGATGTCAAAACCCCTGGCAACCCCG 328
QY 53 TATCCCGATGAGACCTTCAGAAAGTGGGATCCCATCATATAGCACTACTAGGCT 122
Db 329 TATCCCGATGAGACCTTCAGAAAGTGGGATCCCATCATATAGCACTACTAGGCT 388
QY 123 GCGAGTATCATATGTGGTGTCTCATCAAGTGTATCTGGATTAATAACTTCTCT 182
Db 389 GCGAGTATCATATGTGGTGTCTCATCAAGTGTATCTGGATTAATAACTTCTCTCT 448
QY 183 CTGCGGCGAGCTTCCATCTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTG 242
Db 449 CTGCGGCGAGCTTCCACTTCATCCGAGGAGCAGCTGTGTGACGGAGAGCTGGACTG 508
QY 243 TCCCTTGGGGGAGGAGGAGCAGCTGTCAAGAGCTTCCCGAGGCGCTGAGTGGC 302
Db 509 TCCCTTGGGGGAGGAGGAGCAGCTGTGTCAAGAGCTTCCCGAGGCGCTGAGTGGC 568
QY 303 AGTCGCGCTCTCCAAAGGACCGATCCACACTGCGAGTGTGACTCGGCCACAGGGAATG 362
Db 559 AGTCGCGCTCTCCAAAGGACCGATCCACACTGCGAGTGTGACTCGGCCACAGGGAATG 628
QY 363 GTTCTCTGCTGTTTCGACAACTTCAGAGAGCTTCGCTGAGACAGCTCTAGGCAGAT 422
Db 629 GTTCTCTGCTGTTTCGACAACTTCAGAGAGCTTCGCTGAGACAGCTCTAGGCAGAT 688
QY 423 GGGCTTACAGCAACCACTTTTCAGAGCTGTGAGATTGGCCCAAGACAGGATCTGGA 482
Db 689 GGGCTTACAGCAACCACTTCAGAGCTGTGAGATTGGCCCAAGACAGGATCTGGA 748

RESULT 14

US-10-101-510-634
; Sequence 634, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 634

QY 483 TGTGTTGAAATCACAGAAAACAGCCAGGAGCTTGCATGGGAATCAAGTGGGCGCTG 542
Db 749 TGTGTTGAAATCACAGAAAACAGCCAGGAGCTTGCATGGGAATCAAGTGGGCGCTG 808
QY 543 TCTCTAGGCTCCCTGGTCTCCCTGCACTCTCTTGGCTGTGGGAAGAGCTGAGACCCC 602
Db 809 TCTCTAGGCTCCCTGGTCTCCCTGCACTCTCTTGGCTGTGGGAAGAGCTGAGACCCC 868
QY 603 CCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTTGGCAGGTGAGCATCCA 662
Db 869 CCGTGTGGTGGTGGGAGGAGGCTCTGTGGATTCTTGGCCCTTGGCAGGTGAGCATCCA 928
QY 663 GTACGACAAAACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCAGGC 722
Db 929 GTACGACATACAGCAGCTCTGTGGAGGAGCATCTCTGGACCCCACTGGGTCTCAGGC 988
QY 723 AGCCCACTGCTCAGGAAACATACCCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAGA 782
Db 989 AGCCCACTGCTCAGGAAACATACCCGATGTGTTCAACTGGAAGGTGGGCGAGGCTCAGA 1048
QY 783 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATATTGAATTCACCCC 842
Db 1049 CAAACTGGGAGCTTCCCATCCCTGGCTGTGGCAAGATCATCATATTGAATTCACCCC 1108
QY 843 CATGTACCCCAAGACATGACATGCGCCCTCATGAAGTCCAGTTCCTCCTCTCTCTC 902
Db 1109 CATGTACCCCAAGACATGACATGCGCCCTCATGAAGTCCAGTTCCTCCTCTCTCTC 1168
QY 903 AGGCACAGTCAGGCCCATCTGTCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 962
Db 1169 AGGCACAGTCAGGCTCATCTGTCTGCGCTTCTTTGATGAGGAGCTCACTCCAGCCACCCC 1228
QY 963 ACTCTGGATCATTTGGATGGGCTTTTACGAAGAGAAATGGAGGAAGATGTTCACTACT 1022
Db 1229 ACTCTGGATCATTTGGATGGGCTTTTACGAAGAGAAATGGAGGAAGATGTTCACTACT 1288
QY 1023 GCTGCGAGCTCAGTCCAGTCTATTGACAGCACACCGTGTCAATGCAGAGATCGGTACCA 1082
Db 1289 GCTGCGAGCTCAGTCCAGTCTATTGACAGCACACCGTGTCAATGCAGAGATCGGTACCA 1348
QY 1083 GGGGGAAGTCACCGAGAAGATGATGTGTGAGGAGCTCCCGAAGGGGGTGTGACACCTG 1142
Db 1349 GGGGGAAGTCACCGAGAAGATGATGTGTGAGGAGCTCCCGAAGGGGGTGTGACACCTG 1408
QY 1143 CGAGGTGACAGTGTGGGCGCTGATGTACCAATCTGACAGTGGCATGTGTGGGCT 1202
Db 1409 CGAGGTGACAGTGTGGGCGCTGATGTACCAATCTGACAGTGGCATGTGTGGGCT 1468
QY 1203 CGTTAGTGGGCTATGGCTGCGGGGCGCCGAGCACCCAGGAGTATACACCAAGGTCTC 1262
Db 1469 CGTTAGTGGGCTATGGCTGCGGGGCGCCGAGCACCCAGGAGTATACACCAAGGTCTC 1528
QY 1263 AGCCTATCTCAACTGGATCTCAATGTCTGGAAAGGCTGAGCTG 1305
Db 1529 AGCCTATCTCAACTGGATCTCAATGTCTGGAAAGGCTGAGCTG 1571

LENGTH: 2165
TYPE: DNA
ORGANISM: Homo sapiens
US-10-101-510-634

Query Match 99.2%; Score 1294; DB 15; Length 2165;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGATCTGACAGTCACTGAAACAGCCTCGATGTCGAAACCCCTGGGAAACCC 60
DB 310 ATGGATCTGACAGTCACTGAAACAGCCTCGATGTCGAAACCCCTGGGAAACCC 369
QY 61 CATTATCCCATCGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 120
DB 370 CATTATCCCATCGAGACCTTCAGAAAGTGGGATCCCATCATCATAGCACTACTGAGC 429
QY 121 CTGGCAGTATCATCATTTGCTGTTCTCTCATCAAGGTGATTCTGGATAAATACTACTTC 180
DB 430 CTGGCAGTATCATCATTTGCTGTTCTCTCATCAAGGTGATTCTGGATAAATACTACTTC 489
QY 181 CTCTGGGCGAGCTCTCCATCTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGCAC 240
DB 490 CTCTGGGCGAGCTCTCCATCTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGCAC 549
QY 241 TGTCCCTTGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTCGAGTG 300
DB 550 TGTCCCTTGGGAGGAGCAGGAGCACTGTGTCAAGAGCTTCCCGAAGGGCCTCGAGTG 609
QY 301 CGAGTCCGCTCTCCAAAGACCGATCCACACTCGAGGTCTGGAATCGGCCACAGGGAAC 360
DB 610 CGAGTCCGCTCTCCAAAGACCGATCCACACTCGAGGTCTGGAATCGGCCACAGGGAAC 669
QY 361 TGGTCTCTGCTGTTTCACAACTTCACAGAGCTCTCGCTGACAGCCTGTAGGAG 420
DB 670 TGGTCTCTGCTGTTTCACAACTTCACAGAGCTCTCGCTGACAGCCTGTAGGAG 729
QY 421 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCGACAGGATCTG 480
DB 730 ATGGGCTACAGCAGCAAAACCCACTTTTCAGAGCTGTGGAGATTGGCCCGACAGGATCTG 789
QY 481 GATGTTGTTGAATACAGAAACAGCCAGGA-GCTTCGATGCGGAATCAAGTGGGCC 539
DB 790 GATGTTGTTGAATACAGAAACAGCCAGGAGCTTCGATGCGGAATCAAGTGGGCC 849
QY 540 CTGCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGCCTGTGGGAAGAGCCTGAAGAC 599
DB 850 CTGCTCTCAGGCTCCCTGCTCTCCCTGCACTGTCTTGCCTGTGGGAAGAGCCTGAAGAC 909
QY 600 CCCCCTGTGGTGGGAGGAGGCTCTGTGATTTTGGCTTGGCAGTTCAGAT 659
DB 910 CCCCCTGTGGTGGGAGGAGGCTCTGTGATTTTGGCTTGGCAGTTCAGAT 969
QY 660 CCAATACGAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCTCACTGGGCTCTCAC 719
DB 970 CCAATACGAAACAGCAGCTCTGTGGAGGAGCATCTCGACCCCTCACTGGGCTCTCAC 1029
QY 720 GGCAGCCCACTGCTTCAGAAACATACCGATGTGTTCAATCGAAGGTGGGGAGGCTC 779
DB 1030 GGCAGCCCACTGCTTCAGAAACATACCGATGTGTTCAATCGAAGGTGGGGAGGCTC 1089
QY 780 AGCAAACTGGGAGCTTCCCATCCCTGCTGTGGCCAGAGATCATCATTAATTTCAA 839
DB 1090 AGCAAACTGGGAGCTTCCCATCCCTGCTGTGGCCAGAGATCATCATTAATTTCAA 1149
QY 840 CCCCATGTATCCCAAAGACATGACATCGCCCTCATGAAGCTGCAAGTTCCTCACTTT 899
DB 1150 CCCCATGTATCCCAAAGACATGACATCGCCCTCATGAAGCTGCAAGTTCCTCACTTT 1209
QY 900 CTCAGGCAAGTTCAGGCCCCATCTGTCTGCTGCTCTTTGATGAGGAGCTCACTCCAGCCAC 959
DB 1210 CTCAGGCAAGTTCAGGCCCCATCTGTCTGCTGCTCTTTGATGAGGAGCTCACTCCAGCCAC 1269

RESULT 15

US-09-851-588-5
; Sequence 5, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND N
; FILE REFERENCE: A-68829-1/DJB/JJD/AMS
; CURRENT APPLICATION NUMBER: US/09/851,588
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: US 09/642,352
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: US 09/656,002
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-851-588-5

Query Match 99.0%; Score 1292; DB 9; Length 2079;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 3 GGATCTCTGACAGTCACTCTGAAACAGCCTCGATGTCGATGTCGAAACCCCTGGGAAACCCCG 62
DB 216 GGATCTCTGACAGTCACTCTGAAACAGCCTCGATGTCGATGTCGAAACCCCTGGGAAACCCCG 275
QY 63 TATCCCCATCGAGACCTTCAGAAAG-GTGGGATCCCCATCATCATGACTACTGAGCC 121
DB 276 TATCCCCATCGAGACCTTCAGAAAGTGGGATGTCGATGTCGATGTCGAAACCCCTGGGAAACCCCG 335
QY 122 TGGCAGTATCATCATTTGCTGCTCTCATCAAGGTGATTCTGGATAAATACTACTTCC 181
DB 336 TGGCAGTATCATCATTTGCTGCTCTCATCAAGGTGATTCTGGATAAATACTACTTCC 395
QY 182 TTTGCGGAGGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGACT 241
DB 396 TTTGCGGAGGAGCTCTCCACTTCATCCGAGGAGCAGCTGTGTGACGAGAGCTGACT 455
QY 242 GTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCAAGTG 301
DB 456 GTCCCTTGGGAGGAGCAGGAGCAGCTGTGTCAAGAGCTTCCCGAAGGGCCTGCAAGTG 515

302	QY	CAGTCCGCTCTCGAAGACCGATCCACACTGAGAGTGTGGACTGTGGCCACAGGGAAC	361
516	Db	CAGTCCGCTCTCCAAAGACCGATCCACACTGAGAGTGTGGACTGTGGCCACAGGGAAC	575
362	QY	GGTTCTCTGCCTCTTTCCACAACTTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGA	421
576	Db	GGTTCTCTGCCTTTCTCGACAACTTCACAGAAGCTCTCGCTGAGACAGCCTGTAGGCAGA	635
422	QY	TGGGCTACAGCAGCAAAACCACTTCAGAGCTGTGGAGATGGCCCHAGACCGATCTGG	481
636	Db	TGGGCTACAGCAGCAAAACCACTTCAGAGCTGTGGAGATGGCCCHAGACCGAGATCTGG	695
482	QY	ATGTGTGTGAAATTCACAGAAAACGCGAGAGCTTCGCATGCGGAACTCAAGTGGGCGCT	541
696	Db	ATGTGTGTGAAATTCACAGAAAACGCGAGAGCTTCGCATGCGGAACTCAAGTGGGCGCT	755
542	QY	GTCTCTCAGGCTCCGTGTCTCCGTGCATGTCTTGCTGTGGAGAGAGCTGAAAGACC	601
756	Db	GTCTCTCAGGCTCCGTGTCTCCGTGCATGTCTTGCTGTGGAGAGAGCTGAAAGACC	815
602	QY	CCCGTGTGTGGGTGGGGAGAGCCCTCTGTGATTTCTTGGCTTGGCAGGTCAAGATCC	661
816	Db	CCCGTGTGTGGGTGGGGAGAGCCCTCTGTGATTTCTTGGCTTGGCAGGTCAAGATCC	875
662	QY	AGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGTGACCCGCCACTGGGTCTCACGG	721
876	Db	AGTACGACAAACAGCAGCTCTGTGGAGGAGCATCTGTGACCCGCCACTGGGTCTCACGG	935
722	QY	CAGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGGCAGGCTCAG	781
936	Db	CAGCCCACTGCTTCAGGAAACATACCGATGTGTCAACTGGAAGGTGCGGGCAGGCTCAG	995
782	QY	ACAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATTTGAATTCAAAC	841
996	Db	ACAACTGGGCAGCTTCCCATCCCTGGCTGTGGCCAAAGATCATCATCATTTGAATTCAAAC	1055
842	QY	CCATGTACCCAAAGACAATGACATGCCCTCATGAAGCTGCAGTTCGCCACTCACTTTCT	901
1056	Db	CCATGTACCCAAAGACAATGACATGCCCTCATGAAGCTGCAGTTCGCCACTCACTTTCT	1115
902	QY	CAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCAGTCCAGGCCACCC	961
1116	Db	CAGGCACAGTCAGGCCCATCTGTCTGCCCTTCTTTGATGAGGAGCTCAGTCCAGGCCACCC	1175
962	QY	CAGTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAGATGGAGGGAAGATGCTGACATAC	1021
1176	Db	CAGTCTGGATCATTTGGATGGGGCTTTACGAAGCAGAGATGGAGGGAAGATGCTGACATAC	1235
1022	QY	TGCTGCAAGGCTCAGTCCAGGTCATTGACAGCAACGGTGCATGACAGAGATGCGTACC	1081
1236	Db	TGCTGCAAGGCTCAGTCCAGGTCATTGACAGCAACGGTGCATGACAGAGATGCGTACC	1295
1082	QY	AGGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGGAAGGGGTGTGGAACCT	1141
1296	Db	AGGGGAAGTCACCGAGAAGATGATGTGCAGGCATCCCGGAAGGGGTGTGGAACCT	1355
1142	QY	GCACGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCA	1201
1356	Db	GCACGGGTGACAGTGTGGGCCCTGATGTACCAATCTGACCAAGTGGCATGTGGTGGCA	1415
1202	QY	TCGTTAGCTGGGGCTATGGCTGCGGGGGCCGAGCACCCACAGAGTATACCAAGTCT	1261
1416	Db	TCGTTAGCTGGGGCTATGGCTGCGGGGGCCGAGCACCCACAGAGTATACCAAGTCT	1475
1262	QY	CAGCCTATCTCAACTGATCTCAAAATGTGTGGAAGGCTGAGCTG	1305
1476	Db	CAGCCTATCTCAACTGATCTCAAAATGTGTGGAAGGCTGAGCTG	1519

Search completed: June 8, 2004, 03:57:26
Job time : 607 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 59 Seconds
(without alignments)
2083.189 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKKLRP.....VYTKVSAYLNWYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_24Jan04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2342	100.0	435	4 AAY72558	Human ser
2	2342	100.0	435	5 AAG78577	Human ser
3	2342	100.0	435	6 ABU04931	Human exp
4	2342	100.0	435	6 ABU04912	Human exp
5	2338	99.8	435	2 AAY06437	Human pro
6	2338	99.8	435	6 ABU04930	Human exp
7	2337	99.8	437	5 AAG79359	CJA8 pref
8	2337	99.8	437	6 ABU04920	Human exp
9	2337	99.8	437	6 ABU04935	Human exp
10	2337	99.8	437	6 ABU04936	Human exp
11	2337	99.8	437	6 ABU04915	Human exp
12	2337	99.8	437	7 AD880525	Cvarian c
13	2329	99.4	492	4 AAY72559	Human ser
14	2329	99.4	492	6 ABU04913	Human exp
15	2329	99.4	492	6 ABU04932	Human exp
16	2319	99.0	437	4 AAE06931	Human mem
17	2319	99.0	437	6 ABU04917	Human exp
18	2319	99.0	437	6 ABU04922	Human exp
19	2297.5	98.1	432	3 AAY99417	Human PRO
20	2297.5	98.1	432	4 AAB566166	Protein o
21	2297.5	98.1	432	4 AAB29188	Human PRO
22	2297.5	98.1	432	4 AAB387581	Human PRO
23	2297.5	98.1	432	5 AAG95906	Human sec
24	2297.5	98.1	432	5 ABU76635	Tumour-as
25	2297.5	98.1	432	6 ABU58564	Human PRO

26	2297.5	98.1	432	6 ABU88112	Novel hum
27	2297.5	98.1	432	6 ABU84427	Human sec
28	2297.5	98.1	432	6 ABR66301	Human sec
29	2297.5	98.1	432	6 ABR65691	Human sec
30	2297.5	98.1	432	6 ABU99631	Human sec
31	2297.5	98.1	432	6 ABU82870	Human PRO
32	2297.5	98.1	432	6 ABU89991	Novel hum
33	2297.5	98.1	432	6 ABR68240	Human sec
34	2297.5	98.1	432	6 ABU96293	Novel hum
35	2297.5	98.1	432	6 ABU92724	Human sec
36	2297.5	98.1	432	6 ABO08801	Human sec
37	2297.5	98.1	432	6 ABO02853	Human sec
38	2297.5	98.1	432	6 ABR75007	Human sec
39	2297.5	98.1	432	6 ABR94769	Human sec
40	2297.5	98.1	432	6 ABU85742	Human PRO
41	2297.5	98.1	432	6 ABU98902	Novel hum
42	2297.5	98.1	432	6 ABU98117	Novel hum
43	2297.5	98.1	432	6 ABU91823	Novel hum
44	2297.5	98.1	432	6 ABU89516	Human PRO
45	2297.5	98.1	432	6 ABU86357	Human sec

ALIGNMENTS

RESULT 1

AAY72558
ID AAY72558 standard; protein; 435 AA.

XX AAY72558;

XX AC

XX XX

XX 02-MAY-2001 (first entry)

XX XX

XX Human seripancrin protein.

XX XX

XX Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;

XX KW

XX arthritis; chronic obstructive pulmonary disorder; COPD; cancer;

XX KW

XX osteoporosis; aberrant wound healing; angiogenesis; diabetes;

XX KW

XX inflammatory disorder; stroke; cardiovascular disease; gene therapy;

XX KW

XX vaccine; cytostatic; cerebroprotective; vulnary; osteopathic.

XX OS

XX Homo sapiens.

XX XX

XX WO200104141-A2.

XX XX

XX 18-JAN-2001.

XX XX

XX 04-JUL-2000; 2000WO-EP006211.

XX XX

XX 12-JUL-1999; 99EP-00113428.

XX PA

(MERE) MERCK PATENT GMBH.

XX FI

Suendermann B, Hofmann U, Matzku S, Wilbert O;

XX XX

WPI; 2001-147177/15.

XX DR

N-PSDB; AAD02556.

XX XX

New extracellular serine protease Seripancrin, useful for treating

PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,

PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound

PT healing.

XX XX

Claim 2; Page 39-40; 45pp; English.

XX XX

The present invention relates to seripancrin polynucleotides, and

CC polypeptides encoded by them. Seripancrin are members of serine protease

CC family. This protein contains a transmembrane domain, a low density

CC lipoprotein (LDL) domain, protease domain and a scavenger receptor

CC cysteine-rich (SCR) domain. The LDL and SCR domains help to define the

CC specificity of seripancrin's intra and intermolecular interactions. The

CC polynucleotides and polypeptides of the invention are useful for treating

CC and diagnosing diseases such as arthritis, chronic obstructive pulmonary

CC disorder (COPD), cancer, osteoporosis, aberrant wound healing,
CC angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular
CC diseases. Seripancrin genes are useful in chromosome localisation
CC studies, as tools for tissue expression studies and also in gene therapy.
CC The polypeptides of the invention are used for identifying agonists and
CC antagonists useful for treating conditions associated with seripancrin
CC imbalance. These polypeptides are also useful as vaccines. The present
CC sequence is seripancrin protein. The seripancrin gene is located on human
CC chromosome 11q22-q23
XX
XX
SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 4; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKVIDLKYYF 60
Db 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKVIDLKYYF 60
QY 61 LCGQPLHFTPRKQLCDGELDCPLGDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
Db 61 LCGQPLHFTPRKQLCDGELDCPLGDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
QY 121 WFSACFDNFTEALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNNSGP 180
Db 121 WFSACFDNFTEALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNNSGP 180
QY 181 CLSGSLVSLHCLACGKSLKTPRVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHHWLT 240
Db 181 CLSGSLVSLHCLACGKSLKTPRVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHHWLT 240
QY 241 AAHCFRKHDTVFNWVRAGSKLGSPFSLAVAKIIIEFNPMYKNDIALMKLQFPLTF 300
Db 241 AAHCFRKHDTVFNWVRAGSKLGSPFSLAVAKIIIEFNPMYKNDIALMKLQFPLTF 300
QY 301 SGTVPICLPFPDELTPTPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAY 360
Db 301 SGTVPICLPFPDELTPTPLWIIGWFTKQNGKMSDILLQASVQVIDSTRCNADAY 360
QY 361 QGEVTERKMCACIPGGVDTCQDGGPLMVQSDQVHVGVSVGCGGSGTGVYTKV 420
Db 361 QGEVTERKMCACIPGGVDTCQDGGPLMVQSDQVHVGVSVGCGGSGTGVYTKV 420
QY 421 SAYLWNIYNWKAEL 435
Db 421 SAYLWNIYNWKAEL 435

RESULT 2
AAG78577
ID AAG78577 standard; protein; 435 AA.

XX AAG78577;
XX
XX
DT 07-MAY-2002 (first entry)
XX
XX Human serine protease D-G amino acid sequence.

DE Serine protease; D-G; human; zymogen; enzyme; cytostatic;
KW anti-inflammatory; dermatological; anticoagulation; cancer; skin disorder;
KW neuropathic pain; inflammatory disorder; coagulation diathesis;
KW thrombosis; laundry detergent; skin care; gene therapy.
XX
OS Homo sapiens.
XX Key Location/Qualifiers
FH Region 31..52
FT /note= "hydrophobic transmembrane domain"
FT Active-site 202..203
FT /note= "residues spanning the zymogen cleavage site"
FT Active-site 243
FT /note= "catalytic triad residue"

FT Active-site 339 /note= "catalytic triad residue"
FT Misc-difference 361 /note= "Encoded by CTG".
FT Active-site 385 /note= "catalytic triad residue"
XX
XX W0200202011-A1.
XX
XX 10-JAN-2002.
XX
XX 08-JUN-2001; 2001WO-US018568.
XX
XX 30-JUN-2000; 2000US-00607745.
XX
XX (ORTH) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow AL, Qi J, Andrade-Gordon P;
XX
XX MPI; 2002-106601/14.
XX N-PSDB; AAI64284.
XX
XX Nucleic acid encoding a serine protease called D-G protein which is
XX useful for identifying modulators that are useful for treating a
XX condition which is mediated by protease D-G, e.g. cancer, skin disorders,
XX or neuropathic pain.
XX
XX Claim 13; Fig 1B; 81pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid that
XX encodes a serine protease called D-G protein. The activity of the protein
XX of the invention may be described as cytostatic, anti-inflammatory,
XX dermatological and anticoagulation. The serine protease of the invention
XX is a member of the trypsin/chymotrypsin-like (S1) serine protease family,
XX which play an important role in processes such as digestion and
XX regulatory amplification cascades through the proteolytic activation of
XX inactive zymogen precursors. Protease D-G modulating compounds are useful
XX for treating a condition which is mediated by protease D-G, e.g. cancer,
XX skin disorders, neuropathic pain, inflammatory disorders, or coagulation
XX diathesis/thrombosis. The polynucleotide encoding the protease is useful
XX for identifying, detecting or isolating mutant forms of DNA molecules
XX encoding the protease. The protease is useful for identifying modulators
XX of the functional protease. The D-G protein can be used for formulation
XX of compositions for laundry detergents and skin care products. Protease D
XX -G gene therapy may be used to introduce protease D-G into the cells of
XX target organisms. As the D-G protein is derived from a human, it is less
XX likely to produce an allergic reaction in sensitive individuals when used
XX in formulations for laundry detergents and skin care products. The
XX current sequence represents the human serine protease D-G amino acid
XX sequence
SQ Sequence 435 AA;

Query Match 100.0%; Score 2342; DB 5; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKVIDLKYYF 60
Db 1 MDPDSQPLNSLDVPLKRPRIKPRIPMETFRKVGPIPIIIALLSLASIIIVVVLKVIDLKYYF 60
QY 61 LCGQPLHFTPRKQLCDGELDCPLGDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
Db 61 LCGQPLHFTPRKQLCDGELDCPLGDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGN 120
QY 121 WFSACFDNFTEALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNNSGP 180
Db 121 WFSACFDNFTEALAEATACRQMGYSKPTFRFRAVEIGPDQDLVWEITENSQELRMNNSGP 180
QY 181 CLSGSLVSLHCLACGKSLKTPRVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHHWLT 240
Db 181 CLSGSLVSLHCLACGKSLKTPRVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHHWLT 240

QY 241 AAHCFRKHDTVFNWKVVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
Db 241 AAHCFRKHDTVFNWKVVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
Db 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
QY 361 QGEVTERKMCAGIPEGGVDTCQGDSDGGLMYQSDQHWVGVISWGYCGGSPSTPGVYTKV 420
Db 361 QGEVTERKMCAGIPEGGVDTCQGDSDGGLMYQSDQHWVGVISWGYCGGSPSTPGVYTKV 420
QY 421 SAYLNWIYNNWKAEL 435
Db 421 SAYLNWIYNNWKAEL 435
RESULT 3
ID ABU04931 standard; protein; 435 AA.
XX AC ABU04931;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1597.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX FN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US009671.
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0326370P.
XX PR 04-DEC-2001; 2001US-0336780P.
XX PR 20-FEB-2002; 2002US-0358985P.
XX PA (ZYCO-) ZYCO INC.
XX PI Chicx RM, Tomlinson AJ, Urban RG;
XX DR WPI; 2003-040607/03.
XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX PS Example 2; SEQ ID NO 1597; 134pp; English.
XX CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for

CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 435 AA;
SQ Query Match 100.0%; Score 2342; DB 6; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0;
QY 1 MDPDSQPLNSLDVXPLRKPRIPIPRIPMETFRKVGIPITIIALLSLASIIIVVVLKIVLDKYYF 60
Db 1 MDPDSQPLNSLDVXPLRKPRIPIPRIPMETFRKVGIPITIIALLSLASIIIVVVLKIVLDKYYF 60
QY 61 LCGQPLHPIPRKQLCDGELDCPLGDEDEHCVKSFPEGPAVAVRLSKDRSTLOVLSATGN 120
Db 61 LCGQPLHPIPRKQLCDGELDCPLGDEDEHCVKSFPEGPAVAVRLSKDRSTLOVLSATGN 120
QY 121 WFSACFDNFTEALAEATACRQMGYSKPTPRAVEIGPDQDLVVEITENSQELMRNNSGP 180
Db 121 WFSACFDNFTEALAEATACRQMGYSKPTPRAVEIGPDQDLVVEITENSQELMRNNSGP 180
QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDPHWLVT 240
Db 181 CLSGSLVSLHCLACGSKLTPRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDPHWLVT 240
QY 241 AAHCFRKHDTVFNWKVVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
Db 241 AAHCFRKHDTVFNWKVVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
Db 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAY 360
QY 361 QGEVTERKMCAGIPEGGVDTCQGDSDGGLMYQSDQHWVGVISWGYCGGSPSTPGVYTKV 420
Db 361 QGEVTERKMCAGIPEGGVDTCQGDSDGGLMYQSDQHWVGVISWGYCGGSPSTPGVYTKV 420
QY 421 SAYLNWIYNNWKAEL 435
Db 421 SAYLNWIYNNWKAEL 435
RESULT 4
ID ABU04912 standard; protein; 435 AA.
XX AC ABU04912;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1578.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX FN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US009671.
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985P.
PA (ZYCO-) ZYCOS INC.
XX Chiciz RM, Tomlinson AJ, Urban RG;
XX WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX
XX
XX Example 2; SEQ ID NO 1578; 134pp; English.
XX
XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 435 AA;
SQ
Query Match 100.0%; Score 2342; DB 6; Length 435;
Best Local Similarity 100.0%; Pred. No. 2.8e-202; Mismatches 0; Gaps 0;
Matches 435; Conservative 0; Indels 0;
QY 1 MDPSDQPLNSLDVPLKRPIMPETFRKVGPIPIIIALLSLASIIIVVLKIVLDKYF 60
Db 1 MDPSDQPLNSLDVPLKRPIMPETFRKVGPIPIIIALLSLASIIIVVLKIVLDKYF 60
QY 61 LCGQLHPIPRKQCDGLDCLGDEBHCVKSPPEGPAVAVRLSKORSTLQVLD SATGN 120
Db 61 LCGQLHPIPRKQCDGLDCLGDEBHCVKSPPEGPAVAVRLSKORSTLQVLD SATGN 120
QY 121 WFSACFDNFTALATACRQGYSKPTFRAVEIGPDQDLQDLDVITENSQELRMENSSGP 180
Db 121 WFSACFDNFTALATACRQGYSKPTFRAVEIGPDQDLQDLDVITENSQELRMENSSGP 180
QY 181 CLSGSLVSLHCLACKSKLTPRVVGGEASVDSFPWQVSIQYDQHVCGGSILDPHVLT 240
Db 181 CLSGSLVSLHCLACKSKLTPRVVGGEASVDSFPWQVSIQYDQHVCGGSILDPHVLT 240
QY 241 AAHCFKHTDVFNVKVRAGSKLGFSPFLAVAKIIITFENPMYKNDIALMKLQFLTF 300
Db 241 AAHCFKHTDVFNVKVRAGSKLGFSPFLAVAKIIITFENPMYKNDIALMKLQFLTF 300
QY 301 SGTVRPICLPFFDELTATPLWIIGWFTKQNGKMSDILLQASVQVVDSTRCNADAY 360
Db 301 SGTVRPICLPFFDELTATPLWIIGWFTKQNGKMSDILLQASVQVVDSTRCNADAY 360
QY 361 QGEVTEKMKACIPGGVDTCGGSGGLMTYQSDQHVGVVSVGCGGPGSTPGVTKV 420
Db 361 QGEVTEKMKACIPGGVDTCGGSGGLMTYQSDQHVGVVSVGCGGPGSTPGVTKV 420
QY 421 SAYLNWIYNVWKAEI 435
Db 421 SAYLNWIYNVWKAEI 435

RESULT 5
AAY06437
ID AAY06437 standard; protein; 435 AA.
XX
XX AAY06437;
XX 27-SEP-1999 (first entry)
XX Human protease HUPM-6.
XX Serine protease; human; HUPM-6; cell proliferation; cancer;
KW immune disorder; inflammation; therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Modified-site 26
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 59
FT /note= "tyrosine kinase phosphorylation site"
FT Modified-site 93
FT /note= "casein kinase II phosphorylation site"
FT Modified-site 128
FT /note= "N-glycosylated"
FT Modified-site 144
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 148
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 176
FT /note= "N-glycosylated"
FT Modified-site 197
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 200
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 231
FT /note= "casein kinase II phosphorylation site"
FT Active-site 243
FT /note= "characteristic of serine protease"
FT Modified-site 249
FT /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 260
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 303
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 351
FT /note= "protein kinase C phosphorylation site"
FT Modified-site 360
FT /note= "tyrosine kinase phosphorylation site"
FT Modified-site 365
FT /note= "protein kinase C phosphorylation site"
FT Active-site 385
FT /note= "characteristic of serine protease"
XX
XX W09936550-A2.
XX
XX 22-JUL-1999.
XX
XX 12-JAN-1999; 99WO-US000655.
XX
XX 16-JAN-1998; 98US-00008271.
XX
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Hillman JL, Yue H, Guegler KJ, Corley NC, Tang YT;
PI Shah P;
XX WPI; 1999-430616/36.
DR N-PSDB; AAX87154.
XX
XX Novel human protease molecules useful in the treatment of developmental
PT disorders and/or cancers.
PT

PS Claim 1; Page 74-75; 90pp; English.

XX CC The present sequence represents novel human protease HUPM-6, as deduced
CC from the consensus sequence (see AAX87154) of overlapping cDNA clones
CC obtained from various libraries. Northern analysis shows expression of
CC HUPM-6 in gastrointestinal, and male and female reproductive cDNA
CC libraries. Approximately 65% of these libraries are associated with
CC neoplastic disorders, and 22% with the immune response. The invention
CC provides 12 new human proteases, i.e. HUPM-1 to -12 (see AAY06432-43),
CC and the polynucleotides encoding them (see AAX87143-60). Also provided
CC are vectors, host cells and methods for producing HUPM polypeptides, as
CC well as agonists and antagonists of HUPM. Methods for treating or
CC preventing cell proliferative disorders and immune disorders using HUPM
CC or HUPM antagonists are claimed

XX SQ Sequence 435 AA;

Query Match 99.8%; Score 2338; DB 2; Length 435;
Best Local Similarity 99.8%; Pred. No. 6.5e-202;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKVLDKYIF 60
DB 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKVLDKYIF 60
QY 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
DB 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
DB 121 WFSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180
QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
DB 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240
QY 241 AAHCFRKHDTVFNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
DB 241 AAHCFRKHDTVFNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
QY 301 SGTVRPCLPFDEELTPATPLMIIGWFTKONGKMSDILLOASVQVIDSTRCNADAY 360
DB 301 SGTVRPCLPFDEELTPATPLMIIGWFTKONGKMSDILLOASVQVIDSTRCNADAY 360
QY 361 QGEVTEKMWKAGIPGGVDTCQDGGPLMYQSDQVHVGVGCGGSPSTPGVYTKV 420
DB 361 QGEVTEKMWKAGIPGGVDTCQDGGPLMYQSDQVHVGVGCGGSPSTPGVYTKV 420
QY 421 SAYLNWYNWKAEL 435
DB 421 SAYLNWYNWKAEL 435

RESULT 6
ABU04930
ID ABU04930 standard; protein; 435 AA.

XX AC ABU04930;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1596.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX WN WO200278524-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US009671.

XX PR 28-MAR-2001; 2001US-0279495P.

XX PR 21-MAY-2001; 2001US-0292544P.

XX PR 08-AUG-2001; 2001US-0310801P.

XX PR 01-OCT-2001; 2001US-0326370P.

XX PR 04-DEC-2001; 2001US-0326780P.

XX PR 20-FEB-2002; 2002US-0358985P.

XX PA (ZYCO-) ZYCOS INC.

XX PI Chicz RM, Tomlinson AJ, Urban RG;

XX DR WPI; 2003-040607/03.

XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.

XX PS Example 2; SEQ ID NO 1596; 134pp; English.

XX CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC lymphoma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: this sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 435 AA;

Query Match 99.8%; Score 2338; DB 6; Length 435;
Best Local Similarity 99.8%; Pred. No. 6.5e-202;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKVLDKYIF 60

DB 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIALLSLASIIIVVVLKVLDKYIF 60

QY 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120

DB 61 LCGQPLHFIIPRKQCDGELDCPLGDEDEHCHVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120

QY 121 WFSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180

DB 121 WFSACFDNFTALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGP 180

QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240

DB 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSIIDPHWLT 240

QY 241 AAHCFRKHDTVFNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300

DB 241 AAHCFRKHDTVFNKVRAGSDKLGSPFSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300

QY 301 SGTVRPCLPFDEELTPATPLMIIGWFTKONGKMSDILLOASVQVIDSTRCNADAY 360

DB 301 SGTVRPCLPFDEELTPATPLMIIGWFTKONGKMSDILLOASVQVIDSTRCNADAY 360

QY 361 QGEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKV 420
 DB 361 QGEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKV 420
 QY 421 SAYLNWIYNWKAE 435
 DB 421 SAYLNWIYNWKAE 435

RESULT 7

AAG79359

ID AAG79359 standard; protein; 437 AA.

XX AC

XX AAG79359;

XX 21-AUG-2002 (first entry)

XX CJA8 preferred sequence.

XX Colorectal cancer; CGA7; CJA8; modulating protein; screening;

XX drug candidate; vaccine.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

XX Domain

XX 36..63

XX /note= "Transmembrane domain"

XX US2002042067-A1.

XX 11-APR-2002.

XX 08-MAY-2001; 2001US-00851589.

XX 17-AUG-2000; 2000US-00642252.

XX 06-SEP-2000; 2000US-00656002.

XX (MACK/) MACK D.

XX (GISH/) GISH K C.

XX (WILS/) WILSON K E.

XX Mack D, Gish KC, Wilson KE;

XX WPI: 2002-453647/48.

XX N-PSDB; AA72976.

XX Screening drug candidates for treating colorectal cancer, comprises

XX determining the effect of the candidate on the expression profile gene of

XX CGA7 or CJA8.

XX Disclosure; Fig 8; 40pp; English.

XX The sequences given in AAG79356-59 show the colorectal cancer proteins,

XX CGA7 and CJA8. The CGA7 and CJA8 proteins are colorectal cancer

XX modulating proteins and have been mapped to chromosomes 2 (CGA7) and 11

XX (CJA8). These sequences may be used in the method of the invention for

XX screening drug candidates. The method comprises adding a drug candidate

XX to a cell that expresses an expression profile gene encoding CGA7, CJA8

XX or fragments and determining the effect of the drug candidate on the

XX expression of the expression profile gene. The new methods are used to

XX screen bioactive agents for the ability to bind to or modulate the

XX activity of CGA7 or CJA8 and evaluate the effect of a candidate

XX colorectal cancer drug. An antibody to CGA7 or CJA8 can inhibit the

XX activity of CGA7 or CJA8, respectively, and is used to screen for an

XX agent that can interfere with the binding of CGA7 or CJA8 to the

XX antibody. The antibody can be used to treat colorectal cancer. The

XX antibody or a fragment of it is used to localize a therapeutic group to a

XX colorectal cancer tissue, where the therapeutic group is a cytotoxic

XX agent or a radioisotope. Antisense molecules are used to inhibit

XX colorectal cancer in a cell. Nucleic acid segments encoding CGA7 or CJA8

XX are used in a biochip. CGA7, CJA8 or a nucleic acid encoding it are used

XX to elicit an immune response. CGA7 or CJA8 is used to determine the

XX prognosis of an individual with colorectal cancer. Nucleic acid encoding

XX

CC CGA7 or CJA8 can be used in vaccines

XX Sequence 437 AA;

SQ

Query Match

99.8%; Score 2337; DB 5; Length 437;

Best Local Similarity 100.0%; Pred. No. 8e-202;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYFL 61

DB 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYFL 63

QY 62 CGQPLHFIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121

DB 64 CGQPLHFIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDODLDVVEITENSQELMRNSSGPC 181

DB 124 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDODLDVVEITENSQELMRNSSGPC 183

QY 182 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWFPQVSIQYDKQHVCGGSIILDPHWLTA 241

DB 184 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWFPQVSIQYDKQHVCGGSIILDPHWLTA 243

QY 242 AHCFRKHTDVFNKVRAGSDKLSGSPFSLAVAKIIIEFNPMPKNDIALMKLOFLTF 301

DB 244 AHCFRKHTDVFNKVRAGSDKLSGSPFSLAVAKIIIEFNPMPKNDIALMKLOFLTF 303

QY 302 GTVRPCLPFDEBELTPATPLWIIHGFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 361

DB 304 GTVRPCLPFDEBELTPATPLWIIHGFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 363

QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKVS 421

DB 364 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHWVGVISWGYGCGPSTPGVYTKVS 423

QY 422 AYLNWIYNWKAE 435

DB 424 AYLNWIYNWKAE 437

RESULT 8

ABU04920

ID ABU04920 standard; protein; 437 AA.

XX AC

XX ABU04920;

XX 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #1586.

XX Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

XX protease; protease inhibitor; transporter; cytoskeletal protein;

XX receptor; transcription factor; cancer; MHC;

XX major histocompatibility complex; myeloma; colon cancer; gastric cancer;

XX adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

XX WO200278524-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US009671.

XX 28-MAR-2001; 2001US-0279495P.

XX 21-MAY-2001; 2001US-0292544P.

XX 08-AUG-2001; 2001US-0310801P.

XX 01-OCT-2001; 2001US-0326370P.

XX 04-DEC-2001; 2001US-0336780P.

XX 20-FEB-2002; 2002US-0358985P.

XX (ZYCO-) ZYCO INC.

XX PI Chiciz RM, Tomlinson AJ, Urban RG;
XX DR WPI; 2003-040607/03.
XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX PS Example 2; SEQ ID NO 1586; 134pp; English.
XX CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 437 AA;
Query Match 99.8%; Score 2337; DB 6; Length 437;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVVKPLRPIMPETFRKVGPIIIALLSLASIIIVWLKIVLDKYFL 61
DB 4 DPDSQPLNSLDVVKPLRPIMPETFRKVGPIIIALLSLASIIIVWLKIVLDKYFL 63
QY 62 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 64 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACPDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVVEITENSQELMRNNSGPC 181
DB 124 FSACPDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVVEITENSQELMRNNSGPC 183
QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 241
DB 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 243
QY 242 AHCFRKHTDVFNVKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
DB 244 AHCFRKHTDVFNVKVRAGSKLGSFSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
QY 302 GTVRPICLPFDEELTPATPLIIGWGTQKNGKMSDILLQASVOVIDSTECNADDAVQ 361
DB 304 GTVRPICLPFDEELTPATPLIIGWGTQKNGKMSDILLQASVOVIDSTECNADDAVQ 363
QY 362 GEVTEKMKCAGIPEGGVDTCCQDSGGLMYQSDQWHVVGIVSWGCGGSPSTPGYTKVS 421
DB 364 GEVTEKMKCAGIPEGGVDTCCQDSGGLMYQSDQWHVVGIVSWGCGGSPSTPGYTKVS 423
QY 422 AYLNMIYNWKAEL 435
DB 424 AYLNMIYNWKAEL 437
RESULT 9
ID ABU04935 standard; protein; 437 AA.
XX

AC ABU04935;
XX DT 29-JAN-2003 (first entry)
XX DE Human expressed protein tag (EPT) #1601.
XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
KW protease; protease inhibitor; transporter; cytoskeletal protein;
KW receptor; transcription factor; cancer; MHC;
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.
XX OS Homo sapiens.
XX PN WO200278524-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US0009671.
XX PR 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0358985P.
XX (ZYCO-) ZYCOS INC.
XX PA Chiciz RM, Tomlinson AJ, Urban RG;
XX PI WPI; 2003-040607/03.
XX DR New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
XX cytoskeletal proteins, receptors or transcription factors), useful for
XX treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
XX leukemia.
XX PS Example 2; SEQ ID NO 1601; 134pp; English.
XX CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 437 AA;
Query Match 99.8%; Score 2337; DB 6; Length 437;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVVKPLRPIMPETFRKVGPIIIALLSLASIIIVWLKIVLDKYFL 61
DB 4 DPDSQPLNSLDVVKPLRPIMPETFRKVGPIIIALLSLASIIIVWLKIVLDKYFL 63
QY 62 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 64 CGQPLHFIPRKQLCDGELDCPLGDEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACPDNTEALAEACRQMGYSKPTFRAVEIGPDQDLVVVEITENSQELMRNNSGPC 181

Db 124 FSAFDFNFTTEALAEATACRQMGYSKPTFRAVEIGDPDQLDVVEITENSQELRMENSSGPC 183
 Qy 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 243
 Qy 242 AHCFRKHTDVFNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
 Db 244 AHCFRKHTDVFNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
 Qy 302 GTVRPCLPFPDEBELTPATPLMIIGWFTKQNGKMSDILLQASVQVISTRCNADDAAYQ 361
 Db 304 GTVRPCLPFPDEBELTPATPLMIIGWFTKQNGKMSDILLQASVQVISTRCNADDAAYQ 363
 Qy 362 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQVHVGIVSWGCGGSPTPGVYTKVS 421
 Db 364 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQVHVGIVSWGCGGSPTPGVYTKVS 423
 Qy 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 10

ABU04936
 ID ABU04936 standard; protein; 437 AA.

XX AC ABU04936;

XX DT 29-JAN-2003 (first entry)

XX DE Human expressed protein tag (EPT) #1602.

XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX OS Homo sapiens.

XX PN WO200278524-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US009671.

XX PR 28-MAR-2001; 2001US-0279495P.

XX PR 21-MAY-2001; 2001US-0292544P.

XX PR 08-AUG-2001; 2001US-0310801P.

XX PR 01-OCT-2001; 2001US-0328370P.

XX PR 04-DEC-2001; 2001US-0336780P.

XX PR 20-FEB-2002; 2002US-0358985P.

XX PA (ZYCO-) ZYCOS INC.

XX PI Chicx RM, Tomlinson AJ, Urban RG;

XX DR WPI; 2003-040607/03.

XX PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
 PT leukemia.

XX PS Example 2; SEQ ID NO 1602; 134pp; English.

XX CC The invention describes a purified polypeptide, which comprises a
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,
 CC transporter, cytoskeletal protein, receptor or transcription factor. The
 CC polypeptide is useful as an immunogenic composition for eliciting in a
 CC mammal an immunogenic response directed against any of the purified

CC polypeptide. The purified polypeptide, or the antibody that binds to this
 CC polypeptide, is useful for treating cancer. The polypeptide is also
 CC useful for identifying compounds that binds to a naturally processed
 CC class I or class II MHC-binding polypeptide. The polypeptides and
 CC polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukaemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling. Note: this sequence does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 437 AA;

Query Match 99.8%; Score 2337; DB 6; Length 437;
 Best Local Similarity 100.0%; Pred. No. 8e-202;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DPDSQPLNSLDVKPLRKPRIKPRIPMETPRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 61
 Db 4 DPDSQPLNSLDVKPLRKPRIKPRIPMETPRKVGIPPIIIALLSLASIIIVVLIKVLKYYFL 63
 Qy 62 CGOPLHFIIRKQICDGLDCPLGEDDEHCVKSPFEGPFAVRLSKDRSTLQVLDATGNW 121
 Db 64 CGOPLHFIIRKQICDGLDCPLGEDDEHCVKSPFEGPFAVRLSKDRSTLQVLDATGNW 123
 Qy 122 FSACFDNFTEALAEATACRQMGYSKPTFRAVEIGDPDQLDVVEITENSQELRMENSSGPC 181
 Db 124 FSACFDNFTEALAEATACRQMGYSKPTFRAVEIGDPDQLDVVEITENSQELRMENSSGPC 183
 Qy 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
 Db 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 243
 Qy 242 AHCFRKHTDVFNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
 Db 244 AHCFRKHTDVFNKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
 Qy 302 GTVRPCLPFPDEBELTPATPLMIIGWFTKQNGKMSDILLQASVQVISTRCNADDAAYQ 361
 Db 304 GTVRPCLPFPDEBELTPATPLMIIGWFTKQNGKMSDILLQASVQVISTRCNADDAAYQ 363
 Qy 362 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQVHVGIVSWGCGGSPTPGVYTKVS 421
 Db 364 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQVHVGIVSWGCGGSPTPGVYTKVS 423
 Qy 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 11

ABU04915
 ID ABU04915 standard; protein; 437 AA.

XX AC ABU04915;

XX DT 29-JAN-2003 (first entry)

XX DE Human expressed protein tag (EPT) #1581.

XX KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX OS Homo sapiens.

XX PN WO200278524-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US009671.
XX PR 28-MAR-2001; 2001US-0279495P.
XX PR 21-MAY-2001; 2001US-0292544P.
XX PR 08-AUG-2001; 2001US-0310801P.
XX PR 01-OCT-2001; 2001US-0326370P.
XX PR 04-DEC-2001; 2001US-0336780P.
XX PR 20-FEB-2002; 2002US-0358985P.
XX PA (ZYCO-) ZYCOS INC.
XX PI Chicx RM, Tomlinson AJ, Urban RG;
XX DR WPI; 2003-040607/03.
XX XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX XX Example 2; SEQ ID NO 1581; 134pp; English.
XX XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX XX Sequence 437 AA;

Db 364 GEVTERKMCAGIPEGGVDTCCGDSGGPLMVQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 423
Qy 422 AYLNWYNNWKAEL 435
Db 424 AYLNWYNNWKAEL 437
RESULT 12
ADB80525 ID ADB80525 standard; protein; 437 AA.
XX AC ADB80525;
XX DT 04-DEC-2003 (first entry)
XX DE Ovarian cancer-associated protein #45.
XX KW cytostatic; gene therapy; vaccine; ovarian cancer; diagnosis;
XX KW post-operative chemotherapy; radiation therapy; tumour prognosis;
XX KW pre-cancerous lesion detection.
XX OS Homo sapiens.
XX PN W02002102235-A2.
XX PD 27-DEC-2002.
XX PF 18-JUN-2002; 2002WO-US019297.
XX PR 18-JUN-2001; 2001US-0299234P.
XX PR 27-AUG-2001; 2001US-0315287P.
XX PR 05-SEP-2001; 2001US-0317544P.
XX PR 13-NOV-2001; 2001US-0350666P.
XX PR 12-APR-2002; 2002US-0372246P.
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX PI Mack DH, Gish KC;
XX XX WPI; 2003-167431/16.
XX DR N-PSDB; ADB80524.
XX XX Detecting an ovarian cancer-associated transcript in a cell from a
PT patient, comprises contacting a biological sample from the patient with a
PT polynucleotide that hybridizes to an ovarian cancer gene.
XX PS Claim 13; Page 305; 332pp; English.
XX XX The invention relates to a method of detecting an ovarian cancer-
CC associated transcript in a cell from a patient, by contacting a
CC biological sample from the patient with a polynucleotide that selectively
CC hybridizes to a sequence at least 80% identical to any of one of 80
CC nucleic acid sequences given in the specification. The method is useful
CC in diagnosing ovarian cancer and in identifying and using agents and/or
CC targets that inhibit ovarian cancer. The nucleic acid molecule,
CC polypeptide and the antibody may also be used in detecting ovarian
CC cancers, monitoring and early detection of relapse following treatment,
CC monitoring response to therapy, selecting patients for post-operative
CC chemotherapy or radiation therapy, in selecting mode of therapy,
CC determining tumour prognosis, early detection of pre-cancerous lesions,
CC and as vaccines. This sequence corresponds to one of the proteins used
CC for the detection method of the invention.
XX SQ Sequence 437 AA;
Query Match 99.8%; Score 2337; DB 7; Length 437;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 DPDSQPLNSLDVKPLRPKIPMETFRKVGPIIIALLSLASIIIVVLIKVILDKYFL 61
Db 4 DPDSQPLNSLDVKPLRPKIPMETFRKVGPIIIALLSLASIIIVVLIKVILDKYFL 63

QY 62 CQQLHFIIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLD SATGNW 121
 DB 64 CQQLHFIIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFDNFTEALAEATACRMQGYSSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPGC 181
 DB 124 FSACFDNFTEALAEATACRMQGYSSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPGC 183
 QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 241
 DB 184 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWLTA 243
 QY 242 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 301
 DB 244 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 303
 QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVIDSTRCNADDA 361
 DB 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVIDSTRCNADDA 363
 QY 362 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGYCGGPGSTPGVYTKV 421
 DB 364 GEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGYCGGPGSTPGVYTKV 423
 QY 422 AYLNWYNNWK 435
 DB 424 AYLNWYNNWK 437

RESULT 13

AAV72559
 ID AAY72559 standard; protein; 492 AA.

XX AAY72559;
 AC AAY72559;

DT 02-MAY-2001 (first entry)

XX Human seripancrin variant #1 protein.

DE Human; seripancrin; serine protease; chromosome 11q22-q23; therapy;
 XX arthritis; chronic obstructive pulmonary disorder; COPD; cancer;
 KW osteoporosis; aberrant wound healing; angiogenesis; diabetes;
 KW inflammatory disorder; stroke; cardiovascular disease; gene therapy;
 KW vaccine; cytostatic; cerebroprotective; vulnery; osteopathic.

XX Homo sapiens.

OS WO200104141-A2.

PN 18-JAN-2001.

XX 04-JUL-2000; 2000WO-EP006211.

XX 12-JUL-1999; 99EP-00113428.

XX (MERE) MERCK PATENT GMBH.

XX Suendermann B, Hofmann U, Matzku S, Wilbert O;

XX WPI; 2001-147177/15.

DR N-PSDB; AAD02557.

XX New extracellular serine protease Seripancrin, useful for treating
 PT cancer, osteoporosis, arthritis, chronic obstructive pulmonary disease,
 PT diabetes, inflammatory disorders, stroke, angiogenesis and aberrant wound
 PT healing.

XX Disclosure; Page 43-44; 45pp; English.

XX The present invention relates to seripancrin polynucleotides, and
 CC polypeptides encoded by them. Seripancrin are members of serine protease
 CC family. This protein contains a transmembrane domain, a low density
 CC lipoprotein (LDL) domain, protease domain and a scavenger receptor

CC cysteine-rich (SRCR) domain. The LDL and SRCR domains help to define the
 CC specificity of seripancrin's intra and intermolecular interactions. The
 CC polynucleotides and polypeptides of the invention are useful for treating
 CC and diagnosing diseases such as arthritis, chronic obstructive pulmonary
 CC disorder (COPD), cancer, osteoporosis, aberrant wound healing,
 CC angiogenesis, inflammatory disorders, diabetes, stroke and cardiovascular
 CC diseases. Seripancrin genes are useful in chromosome localisation
 CC studies, as tools for tissue expression studies and also in gene therapy.
 CC The polypeptides of the invention are used for identifying agonists and
 CC antagonists useful for treating conditions associated with seripancrin
 CC imbalance. These polypeptides are also useful as vaccines. The present
 CC sequence is seripancrin variant #1 protein. The seripancrin gene is
 CC located on human chromosome 11q22-q23
 XX
 SQ Sequence 492 AA;

Query Match 99.4%; Score 2329; DB 4; Length 492;
 Best Local Similarity 100.0%; Pred. No. 4.9e-201;
 Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLRPMPMETFRKVGPIIIIIALLSLASIIIVVLIKVLDKYYF 60
 DB 1 MDPDSQPLNSLDVKPLRPMPMETFRKVGPIIIIIALLSLASIIIVVLIKVLDKYYF 60
 QY 61 LCGQPLHFIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLD SATGN 120
 DB 61 LCGQPLHFIPRKOLCDGELDCPLGEDEEHCVKSFPEGPVAVRLSKDRSTLQVLD SATGN 120
 QY 121 WFSACFDNFTEALAEATACRMQGYSSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPG 180
 DB 121 WFSACFDNFTEALAEATACRMQGYSSKPTFRAVEIGDPQDLVDVEITENSQELRMNSGPG 180
 QY 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWL 240
 DB 181 CLSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPMQVSIQYDKQHVCGSILDPHWL 240
 QY 241 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
 DB 241 AHCFRKHTDFVFNKVRAGSKLGSFSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTF 300
 QY 301 SGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVIDSTRCNADDA 360
 DB 301 SGTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKXSDILLOASVQVIDSTRCNADDA 360
 QY 361 QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGYCGGPGSTPGVYTKV 420
 DB 361 QGEVTEKMMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGYCGGPGSTPGVYTKV 420
 QY 421 SAYLNWYNNWK 432
 DB 421 SAYLNWYNNWK 432

RESULT 14

ABU04913
 ID ABU04913 standard; protein; 492 AA.

XX AC ABU04913;

DT 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #1579.

DE Translational profiling; expressed protein tag; EPT; kinase; phosphatase;
 KW protease; protease inhibitor; transporter; cytoskeletal protein;
 KW receptor; transcription factor; cancer; cancer; WHC;
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX Homo sapiens.

XX WO200278524-A2.

```
PD XX 10-OCT-2002.
PF XX 28-MAR-2002; 2002WO-US009671.
XX XX
PR 28-MAR-2001; 2001US-0279495P.
PR 21-MAY-2001; 2001US-0292544P.
PR 08-AUG-2001; 2001US-0310801P.
PR 01-OCT-2001; 2001US-0326370P.
PR 04-DEC-2001; 2001US-0336780P.
PR 20-FEB-2002; 2002US-0359895P.
XX XX
PA (ZYCO-) ZYCOS INC.
XX XX
PI Chicx RM, Tomlinson AJ, Urban RG;
XX XX
DR WPI; 2003-040607/03.
XX XX
PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or
PT leukemia.
XX XX
PS Example 2; SEQ ID NO 1579; 134pp; English.
XX XX
CC The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor. The
CC polypeptide is useful as an immunogenic composition for eliciting in a
CC mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to this
CC polypeptide, is useful for treating cancer. The polypeptide is also
CC useful for identifying compounds that binds to a naturally processed
CC class I or class II MHC-binding polypeptide. The polypeptides and
CC polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling. Note: This sequence does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX XX
SQ Sequence 492 AA;

Query Match          99.4%; Score 2329; DB 6; Length 492;
Best Local Similarity 100.0%; Pred. No. 4.9e-201; Mismatches 0; Indels 0; Gaps 0;
Matches 432; Conservative 0;

QY 1 MDPSDQPLNSLDVVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60
DB 1 MDPSDQPLNSLDVVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60
QY 61 LCGQPLHPIPRKOLCDGELDCEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGN 120
DB 61 LCGQPLHPIPRKOLCDGELDCEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGN 120
QY 121 WFSACFDNFTALAEACRQMGYSKPTFRVAVETGPDQDLVDWETENSQBLRNRSSGP 180
DB 121 WFSACFDNFTALAEACRQMGYSKPTFRVAVETGPDQDLVDWETENSQBLRNRSSGP 180
QY 181 CLSGSLVSLHCLACKSLKTRVVGGEASVDSFWQVSIQYDKHQVCGGSLDPHVLVT 240
DB 181 CLSGSLVSLHCLACKSLKTRVVGGEASVDSFWQVSIQYDKHQVCGGSLDPHVLVT 240
QY 241 AAHCFRKHDTVFNKVRAGSKLGSFPPLSAVAKIIITFENPMYKPDNDIALMKLQFPLTF 300
DB 241 AAHCFRKHDTVFNKVRAGSKLGSFPPLSAVAKIIITFENPMYKPDNDIALMKLQFPLTF 300
QY 301 SGTVRPILCLPFFDEBLTPATPLWIGWFTKONGGKMSDILLQASVQVIDSTRCNADDAY 360
DB 301 SGTVRPILCLPFFDEBLTPATPLWIGWFTKONGGKMSDILLQASVQVIDSTRCNADDAY 360
QY 361 QGEVTEKMCAGIPEGGVDTCQDSDGGLMYQSDQWHVGVISWGYCGGPGSTPGVYTKV 420
```

Query Match 99.4%; Score 2329; DB 6; Length 492;

Best Local Similarity 100.0%; Pred. No. 4.9e-201;		Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVVLKIVILDKYIF	60
Db	1	MDPDSQPLNSLDVKPLRKPRIPMETFRKVGIPPIIIALLSLASIIIVVVLKIVILDKYIF	60
Qy	61	LCGQPLHFIIPRQLCDGELDPLGDEBEHCVKSPPEGPAVAVRLSKORSTLOVILDSATGN	120
Db	61	LCGQPLHFIIPRQLCDGELDPLGDEBEHCVKSPPEGPAVAVRLSKORSTLOVILDSATGN	120
Qy	121	WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDODLDWEITENSQELRVNSSGP	180
Db	121	WFSACFDNFTALAEATACRQMGYSKPTFRAVEIGPDODLDWEITENSQELRVNSSGP	180
Qy	181	CLSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLT	240
Db	181	CLSGSLVSLHCLACGKSLKTPRVVGEEASVDSWPQVSIQYDKQHVCGGSIILDPHWLT	240
Qy	241	AAHCFRKHTDVENWVRAGSKLGSFPLAVAKIIIEFNPMYKNDNDIALMKLOEPLTF	300
Db	241	AAHCFRKHTDVENWVRAGSKLGSFPLAVAKIIIEFNPMYKNDNDIALMKLOEPLTF	300
Qy	301	SGTVRPICLPFFDEBELTPATPLIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY	360
Db	301	SGTVRPICLPFFDEBELTPATPLIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAY	360
Qy	361	QGEVTEKMKCAGIPGGVDTCQDSDGGPLMYQSDQHHVGVISWYCGCGPSTPGVYTKV	420
Db	361	QGEVTEKMKCAGIPGGVDTCQDSDGGPLMYQSDQHHVGVISWYCGCGPSTPGVYTKV	420
Qy	421	SAYLNWYVNVK 432	
Db	421	SAYLNWYVNVK 432	

Search completed: June 7, 2004, 08:40:34
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 20 Seconds.
(without alignments)
2092.166 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVKPLRKP.....VYTKVSAYLNWYNVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	584.5	25.0	417	1	S00845	hepsin (EC 3.4.21.
2	582.5	24.9	1019	1	A56318	enteropeptidase (E
3	571	24.4	1035	1	A43090	enteropeptidase (E
4	568.5	24.3	416	1	S33777	hepsin (EC 3.4.21.
5	552	23.6	1034	1	A53663	enteropeptidase (E
6	511	21.8	638	1	KQ3TPL	plasma kallikrein
7	508	21.7	638	1	KQMSPL	plasma kallikrein
8	505	21.6	436	2	JX0172	acrosin (EC 3.4.21
9	504.5	21.5	1524	2	T30337	polyprotein - Afri
10	503	21.5	418	2	A37344	acrosin (EC 3.4.21
11	501.5	21.4	437	2	S19407	acrosin (EC 3.4.21
12	493.5	21.1	421	2	S29599	acrosin (EC 3.4.21
13	491.5	21.0	638	1	KQHUP	plasma kallikrein
14	488.5	20.9	1113	2	JE0315	low-density lipopr
15	481	20.5	421	1	S11674	acrosin (EC 3.4.21
16	478.5	20.4	415	1	A34170	acrosin (EC 3.4.21
17	474	20.2	855	2	JC7731	membrane-bound arg
18	470	20.1	761	2	JCS759	brain-specific ser
19	468.5	20.0	431	2	S47538	acrosin (EC 3.4.21
20	461.5	19.7	277	2	S35340	trypsin (EC 3.4.21
21	460.5	19.7	263	1	KY3TB	chymotrypsin (EC 3
22	458.5	19.6	267	2	S40006	trypsin (EC 3.4.21
23	458.5	19.6	790	1	PLPG	plasmin (EC 3.4.21
24	457.5	19.5	625	1	KFHU1	coagulation factor
25	455.5	19.4	812	1	PLMS	plasmin (EC 3.4.21
26	449	19.2	420	2	A55283	acrosin (EC 3.4.21
27	448	19.1	810	2	I46260	plasmin (EC 3.4.21
28	447.5	19.1	343	1	A57014	proctasin (EC 3.4.
29	445.5	19.0	455	2	A61545	plasmin (EC 3.4.21

ALIGNMENTS

RESULT 1

S00845
hepsin (EC 3.4.21.-) - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C:Accession: S00845
R:Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dome
A:Reference number: S00845; MUID:88209431; PMID:2835076
A:Accession: S00845
A:Molecule type: mRNA
A:Residues: 1-417 <LEY>
A:Cross-references: EMBL:X07732; NID:G32063; PIDN:CAA30558.1; PID:G32064
C:Genetics:
A:Gene: GDB:HPN; TMRSS1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TN>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,322-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match 25.0%; Score 584.5; DB 1; Length 417;
Best Local Similarity 31.5%; Pred.No. 7.5e-41;
Matches 140; Conservative 60; Mismatches 161; Indels 83; Gaps 11;

QY	19	KPRIPMETFRKVGPIIIALLSLASIIIVVLIKVLKDYFLCGQLHPKQLCDGE	78
DB	15	RPKVAULT--AGTLLELTAGASWAIVALLR-----	45
QY	79	LDCLGDEDEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNWFSACDFNTEALATAC	138
DB	46	-----SDQE-----PLYPQVSSADARLMVFKTEGTWRLLCSSRNARVAGLSC	90
QY	139	ROMGYSKPTFRFAVEICPDQDLQDVEITEN-----SOELMRNSSGPECLS	183
DB	91	EEMGF-----LRAL-----THSELVDVTRTAGNTSGFCVDEGRLPHTORLLEVISVCDPR	142
QY	184	GSLSVSLHCLACG-KSLKTPRVVGGEASVDSPWQVSIQYDKQHCVGSLDHPHWLTA	242
DB	143	GRFLAALICQDCGRKLPVDRIVGGRTSLGRWPQVSLRYDGAHLCCGSLSGDWLTA	202
QY	243	HCPRKHTDVN-MKVRAGSKLGSFPPLAVAKIIIEFNWYP-----KNDIALMK	293
DB	203	HCFFPERNVL-SRWKVFAGAAQSPHGLQGVQAVVTHGGYLPFRDPNSSENSDIALVH	262
QY	294	LQFPLTSGTVRTICLPFFDEELTPATPLMTIGNGFTKQNGKMSDILLOASVQVIDSTR	353
DB	263	LSSPLPTEITQPCVLPAAQALVDGKICVTGNGT-QYVGGQAGVLQEARVPIISNDV	321

```

QY 354 CNADYAOQEVETKMKACIGPEGVDTCQDGGFLMYQ-----SDQHVHVGIVSWGYGC 408
DB 322 CNGADFYGNQIKPKMFCAPGEGGIDACQDGGCGFVCEDSISRTPRWRLCGIVSWGTGC 381
QY 409 GGPSTPGVYTKVSAYLWYVWAK 432
DB 382 ALAQKPGVYTKVSDPFWIFQAIK 405

RESULT 2
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C:Accession: A56318; B43090
R:Kitamoto, Y.; Veille, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A:Reference number: A56318; MUID:59234679; PMID:7718557
A:Accession: A56318
A:Molecule type: mRNA
A:Residues: 1-1019 <KIT>
A:Cross-references: GB:U09860; NID:G746412; PIDN:AAC50138.1; PID:G746413
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: B43090
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 749-1019 <KIT>
A:Cross-references: GB:U09860
C:Comment: The mechanism of association with the membrane of the intestinal brush border
ated below) or with amino-terminal myristoylation of the heavy chain.
C:Genetics:
A:Gene: GDB:PRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
A:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
ducts.
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
A:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rep
C:Keywords: Glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:526-631/Domain: C1r/C1s repeat homology <C1r>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:785-1014/Domain: trypsin homology <TRY>
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-836,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 24.9%; Score 582.5; DB 1; Length 1019;
Best Local Similarity 33.8%; Pred. No. 3,3e-40;
Matches 135; Conservative 62; Mismatches 132; Indels 71; Gaps 15;

QY 67 HF-----IPKQLGDCGELDCPLGEDEHCVKSPFEGPAPAVRUSKDRSTLQVLDLSATG 119
DB 647 HFQKNGECVPLVNLCDGHLHCEGSDGDCVRF-----NGRTN 686
QY 120 N-----WFSNCFNFTFALAEACROMGY-----SSKPTFAVEIGPDQDLWVE 164
DB 687 NGLVRFRIQSIWHTACAEWTTQISNDVCOLLGLGSGNSKPIF-STDGPGFVFKLNTAP 745

165 -----ITNSOELMRNSSGPGCLSGSLVSLHC--LACGSLK-----TPRVVGGEEASVDGM 214
746 DGHLLITPSSQ-----CLQDSILRLQCNHKSCKGLAAQDITPKILVGGSSNAKEGAW 796
215 PWQVSIQYDKQHQVCGGSIIDPHWLVTAHHC-FKHHTDVFENWVKVRAGSDKLGGFPFSLAVAK 273
797 PWVGLYGGRLCGASLVSSDLVSAACHVYGRNLEPSKWTAILGLHMKSNLTSPTVP 856
274 III--TEFNPMY---PKNDIALMKLQPLFTSGTVRPICLFFDEELTPATLWITGNG 328
857 RLIDEIVINPHYNNRRKNDIANMHLFEKVNYYDYTPICLPEENQVFPFGNCSLAGWG 916
329 FTKQNGKMSDILQASVOVIDSTRCNADDAVQGVETKMKACIGPEGVDTCQDGGSGP 388
917 -TVYQGTNTIILQEDVPLLSNERCO-QOMPYNITENWICAGYEGGIDSCQDGGSGP 974
389 LMYQ-SDQHVHVGIVSWGYCGGPGSTPGVYTKVSAYLWMI 427
975 LMQCENRWFLAGVTSFGYKCALPNERPGVYVRSRFTWMI 1014

RESULT 3
A43090
enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine
N:Alternate names: enterokinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Apr-2003
C:Accession: A43090; A48874; A61436
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: A43090
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1035 <KIT>
A:Cross-references: GB:U09859; NID:G746410; PIDN:AAB40026.1; PID:G746411
R:LaValle, E.R.; Rehmentulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.; I
J. Biol. Chem. 268, 23311-23317, 1993
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of b
A:Reference number: A48874; MUID:94043122; PMID:8226855
A:Accession: A48874
A:Molecule type: mRNA
A:Residues: 801-1035 <LAV>
A:Cross-references: GB:U19663; NID:G416131; PIDN:AAA16035.1; PID:G416132
A:Note: parts of this sequence, including the amino end of the mature protein, were confi
R:Light, A.; Janaka, H.
J. Protein Chem. 10, 475-480, 1991
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A:Reference number: A61436; MUID:92189715; PMID:1799406
A:Accession: A61436
A:Molecule type: protein
A:Residues: 801-807,'Y',809-827 <LIG>
C:Comment: The mechanism of association with the membrane of the intestinal brush border
embrane attachment using a signal-anchor sequence.
C:Complex: Conversion from membrane-bound to soluble forms may involve further processing
C:Comment: mature enteropeptidase is variously reported to contain two (heavy and light)
fide linked
C:Function:
A:Description: cleaves propeptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
A:Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rep
C:Keywords: Glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein;
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:358-520/Domain: MAM homology <MAM>
F:542-647/Domain: C1r/C1s repeat homology <C1r>
F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>

```

```
QY
134 AETACRQWGY-----SSKPTFRAVEIGPDQLDVVETTENSCELRRKNSSGFCLSGLVSL 189
```

Qy	676
Db	676

LC D G E L D C F L G E D E H C V
| | | | | | | | | | | | | |
LC D G F S H C K D G S D E A H C V

ER-FLNGTANN SGLVQFR--

LQVLDSATGNWFSACFDNE
 :|: | :||:|:
 IQSI-----WHTACAENWW

LEAL 133
|
TTQT 726

Db 727 SDDVCLLGLGTGNSMPPFSGG-GP-----FVKLNTAPNGSLILITASEQCFEDSLILL 780
QY 190 HC--LACGKSL-----KTRPVGGEASVDSWPQVSIQYDKQHVCGGSLDPHWLTAH 243
Db 781 QCNHKSCKGKQVAQVSPKIVGNDREGAWPVVALYNGQLLGCASLVSRSMLVSAH 840
QY 244 C-FRKHDTVFNKVRAGSKLGSF--PSLAVAKIIIIIFENMY-----PKNDIALMKLOPP 297
Db 841 CVYGRNLEPSKKAIIGLHMTSNLTSPOIVTLRDEIVINPHNRRRSDSIAMHLEFK 900
QY 298 LTFSGTVRPLCPFFDEELTPALWIIIGWTFKQNGKMSDILLQASVOVIDSPRCNAD 357
Db 901 VNYTDYIQLPPEENQVPPGRICSIAGKGVYQGSF-ADLQEAADVPLLSNEKCO-Q 958
QY 358 DAYQGEVTEKMKVAGIPEGVTCQDGGPML-YQSDQHWVGVIVSGYCGGPGSTPGV 416
Db 959 QMPENYNTENMKVAGYEGGIDSCQDGGPMLCLENRWLLAGVTSFGYQCALPNRPGV 1018
QY 417 YTKVSAYLNWI 427
Db 1019 YARVPKFTWI 1029
RESULT 6
KORTPL
Plasma kallikrein (EC 3.4.21.34) precursor - rat
N/Alternate names: Fletcher factor; kininogenin; serum kallikrein
C/Species: Rattus norvegicus (Norway rat)
C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999
C/Accession: A3180; S06851; I53041; S06852
R/Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N.G.
Biochemistry 30, 1628-1635, 1991
A/Title: Gene structure and chromosomal localization of plasma kallikrein.
A/Reference number: A39180; PMID:91129236; PMID:1993180
A/Accession: A39180
A/Molecule type: DNA
A/Residues: 1-638 <BEA>
A/Cross-references: GB:J05315
A/Note: The authors translated the codon GAG for residue 81 as Gln
R/Seidah, N.G.; Ladeheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazur
DNA 8, 563-574, 1989
A/Title: The cDNA structure of rat plasma kallikrein.
A/Reference number: A33320; PMID:90091743; PMID:2598771
A/Accession: A33320
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-638 <SEI>
A/Cross-references: GB:M50282; NID:G205010; PIDN:AAA41463.1; PID:G205011
A/Note: Part of this sequence, including the amino ends of both the heavy and light chain
R/Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1989
A/Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A/Reference number: S06851; PMID:90089457; PMID:2597701
A/Accession: S06851
A/Molecule type: protein
A/Residues: 20-45; 391-413 <PQA>
R/Seidah, N.G.; Ladeheim, R.; Mbikay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazur
DNA Cell Biol. 8, 563-574, 1989
A/Title: The cDNA structure of rat plasma kallikrein.
A/Reference number: I53041
A/Accession: I53041
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-638 <RES>
A/Cross-references: GB:M58590; NID:G206721; PIDN:AAA42069.1; PID:G206722
C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
A/Note: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
are linked by one or more disulfide bonds.
C/Genetics:
A/Genes: FK
C/Superfamily: coagulation factor XI; trypsin homology
C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F:391-621/Domain: trypsin homology <TRY>
F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-
F:127,215,308,453,459,494/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:396/Binding site: carboxydrate (Asp) (covalent) #status experimental
F:434,483,578/Active site: His, Asp, Ser #status predicted
Query Match 21.8%; Score 511; DB 1; Length 638;
Best Local Similarity 32.0%; Pred. No. 1.7e-34;
Matches 128; Conservative 62; Mismatches 106; Indels 104; Gaps 17;
QY 64 QPLHFIPRKOLCDG-----ELDCPL--GED--EEHCVKASF-----PEGP 98
Db 290 EPCHF---KIYGVAFEGEELNATFVGADACQETCTKIRCOFFYSLLPQDCKAEGC 345
QY 99 AVAVRLSKDRSTLOVLDATGNWFSACFNFTALAEATACRMGYSSKPTFFRAVEIGPDQ 158
Db 346 KCSLRSLTDSGSPTRITYEAQS-----SGYS-----371
QY 159 DLDVVEITENSQELRMNSGPGCLSGSLVSLHCLAGKSLKTPRVYGGEEASVDSMEPQV 218
Db 372 -LRCKVVESSD-----CTTKINA-RIVGGTNSSLGEWPQV 406
QY 219 SIQ---YDKQHVCGGSLDPHWLTAHCFR--KHTDVFNWVKVRAG-----SKLGSFPPL 269
Db 407 SLQVKLVSNQNMCGGSIIRQWILTAAHCFDGPYPDV--WRIYGGILNLSITNKTTPFS 464
QY 270 AVAKIIIEENPMYKNDIALMKLOFLPTEGTVPEICLPFFDEBELTPALWIIIGWCF 329
Db 465 SIKELIHQKYMSEGSYDALIKLOPLNYTFEQRPICLPKADNTIYTCNWTGWGY 524
QY 330 TKQNGKMSDILLQASVOVIDSTRCNADDAYQGEV--TEKMCAGIPEGVTCQDGGSGGP 388
Db 525 TKER-GETQNILKATIPLPVNEEC--QKKYRDYVITKQMICAGYKEGGIDACKGSDSGP 581
QY 389 LMYQ-SDQHWVGVIVSGYCGGPGSTPGVYTKVSAYLNWI 427
Db 582 LVCKHSGRWLVGITSWGEGCARKEQPGVYTKVAEYIDWI 621
RESULT 7
KOMSPPL
Plasma kallikrein (EC 3.4.21.34) precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 18-Jun-1999
C/Accession: A36557
R/Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemont
DNA Cell Biol. 9, 737-748, 1990
A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compariso
A/Reference number: A36557; PMID:9190844; PMID:2264928
A/Accession: A36557
A/Molecule type: mRNA
A/Residues: 1-638 <SEI>
A/Cross-references: GB:M58588; NID:G200358; PIDN:AAA63393.1; PID:G200359
A/Note: part of this sequence, including the amino ends of both the heavy and light chain
C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
are linked by one or more disulfide bonds.
C/Superfamily: coagulation factor XI; trypsin homology
C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>
F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F:391-621/Domain: trypsin homology <TRY>

F:21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F:127,215,308,396,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 21.7%; Score 508; DB 1; Length 638;
Best Local Similarity 41.0%; Pred. No. 38-34; Indels 34; Gaps 10;
Matches 112; Conservative 45; Mismatches 82; Indels 34; Gaps 10;

QY 174 MRNSQFCGLSGSLVSLHCLACQSLKTP-----RVVGGEEASVDSPWQVSIQ---Y 222
DB 364 MQGSSG-----YSLRLKLVDPDCTTKINARIIVGGINASLGEPWQVSIQVQLV 413
QY 223 DKQHVCGGSLDPHVLTAHGER--KHTDVENKVRAG-----SKLGSFPLAVAKII 276
DB 414 SOTHLCGGSLIGRQWLTAAHCFDGPYPDV--WRIYGSILSLSEITKTPSRKELII 471
QY 277 IEFNPMYPKNDIALMKLOFPLTFSTGVRPCLPFFDEELTPATPLWIIIGWFTKONGK 336
DB 472 HOEYKVSSEGYDIALIKLOPLNLYTEFQKPLPSKADNTIYTCWVTGWYTKEQ-GE 530
QY 337 MSDILLOASVQVIDSTRCNADDAQOGEVTEK-MWCAGIPEGVDTCQDGGSGPLMYQ-SD 394
DB 531 TONILQKATIPVNPBEC--QKKYRDYVINKQMICAGYKEGTDACKGDSGGLVCKHSG 588
QY 395 QHVVGVISWGVGCGGSPSTPGVYTKVSAYLNI 427
DB 589 RNQLVGITSWGCGGCKDQPGVYTKVSEYMDWI 621

RESULT 8

JX0172
acrosin (EC 3.4.21.10) precursor form 3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Jun-2000
C:Accession: JX0172; JX0138
R:Watanabe, K.; Baba, T.; Kashiwabara, S.; Okamoto, A.; Arai, Y.
J. Biochem. 109, 828-833, 1991
A:Title: Structure and organization of the mouse acrosin gene.
A:Reference number: JX0172; MUID:92041732; PMID:1939002
A:Accession: JX0172
A:Molecule type: DNA
A:Residues: 1-436 <WAT>
A:Cross-references: GB:S66245; NID:9238706; PIDN:AA820293.1; PID:9238707
R:Kashiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
J. Biochem. 108, 785-791, 1990
A:Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its gene
A:Reference number: JX0138; MUID:91185335; PMID:2127931
A:Accession: JX0138
A:Molecule type: mRNA
A:Residues: 4-436 <KAS>
A:Cross-references: GB:D00754; NID:9220322; PIDN:BAAC00651.1; PID:9220323
C:Comment: Acrosin is an acrosomal protease that plays an important role in the initial
e-binding activity.
C:Genetics: 26/2; 95/2; 190/1; 238/3
A:Introns: 26/2; 95/2; 190/1; 238/3
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-42/Product: acrosin light chain #status predicted <ALC>
F:43-321/Product: acrosin heavy chain #status predicted <AHC>
F:43-286/Domain: trypsin homology <TRY>
F:22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:225-155,29-162,74-90,178-247,210-226,237-267/Disulfide bonds: #status predicted
F:89,143,241/Active site: His, Asp, Ser #status predicted
F:235/Binding site: substrate (Asp) #status predicted

Query Match 21.6%; Score 505; DB 2; Length 436;
Best Local Similarity 40.9%; Pred. No. 3.3e-34;
Matches 106; Conservative 37; Mismatches 90; Indels 26; Gaps 7;

QY 194 CG-----KSLKTPRVVGGEEASVDSPWQVSIQY-----DKQHVCGGSLDPHVLTA 242
DB 29 CGLRFRRNSQAGTRIVSGSQAQLGAWPMWVSIQITSHNSRYHACGGSLLNSHWLTA 88

QY 243 HCFRKHDTDFNWK-----VRAGSDKLGSPFSLA--VAKIIIEFNPMYPKNDIALMK 293
DB 89 HCFDNKKKYYDRLVFGAQIEYGRNKFVKEPQOERYVQKIVHEKYNVTEGNDIALK 148
QY 294 LQFPLTFSTGVRPCLPFFDEELTPATP--LWIIGWFTKONGKMSDILLQASVQVIDS 351
DB 149 ITPPVTGNGFPGCCLPHF-KAGPPQIPHTCYVTGWIKEKAPRPSFVLMEARVDIDL 207
QY 352 TRCNADDAQOGEVTEKMCAGIPEGVDTCQDGGSGPLMYQSD--QHVHVGVISWGVGC 408
DB 208 DLSNSTQWYNGRVTSTNVCAGYPEK:DTQDGGSGPLMCRDNDVDSFVVVGVITSWGVGC 267
QY 409 GGPSTPGVYTKVSAYLNI 427
DB 268 ARAKRPGVYTATMDYLDWI 286

RESULT 9

T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
C:Accession: T30337
R:Yang, J.C.; Lindsay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from Xer
A:Reference number: Z20829
A:Accession: T30337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: EMBL:U81290; NID:92981640; PID:92981641; PIDN:AAC24717.1
C:Superfamily: trypsin related polyprotein; trypsin homology

Query Match 21.5%; Score 504.5; DB 2; Length 1524;
Best Local Similarity 44.0%; Pred. No. 1.7e-33;
Matches 109; Conservative 38; Mismatches 86; Indels 15; Gaps 8;

QY 194 CGKSLKT-----PRVVGEEASVDSPWQVSIQYDKOHVCGGSLDPHVLTAHCRKH 248
DB 570 CGMAPMTPKMWLPRIIVGGEASPNSPWQVQIFPLRTFHCAGAIISQWILTAAHCIRAA 629
QY 249 TDVFNKVRAGS-DKLGSPFSLAVAKIIIEFNPMYPK--DNDIALMKLOFPLTFSGTV 304
DB 630 EPSY-WTVIAGDENRMLNESTEQIRNKTIRHDNYSYTDNDIALLYLEEPDLNDV 688
QY 305 RPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAQGEV 364
DB 689 RPVCLPEPEEVLTPASVCVVTGWTAEQDQPALG-LQQLPLDLSIIICNT-SYSGEL 746
QY 365 TEKMCAGIPEG-GVDTCQDGGSGPLMYQS--DQHVHVGVISWGVCGGSPSTPGVYTKVS 421
DB 747 TDHMLCAGFPSSKEKDACQDGGSGPLVCQNEKQF:SIYGLVSWGEGCGRVSKPGVYTKVR 806
QY 422 AYLNWIYN 429
DB 807 LFFTWIQN 814

RESULT 10

A37344
acrosin (EC 3.4.21.10) precursor form 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 22-Jun-1999
C:Accession: A37344
R:Klemm, U.; Maier, W.M.; Tsaousidou, S.; Acham, I.M.; Willison, K.; Engel, W.
Differentiation 42, 160-166, 1990
A:Title: Mouse proacrosin-cDNA sequence, primary structure and postmeiotic expression:
A:Reference number: A37344; MUID:90255839; PMID:2111255
A:Accession: A37344
A:Status: preliminary
A:Molecule type: mRNA

A;Residues: 1-418 <KLE>
 A;Cross-references: GB:X52466; NID:949857; PIDN:CAA36704.1; PID:g49858
 C;Superfamily: acrosin; trypsin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase
 F;39-283/Domain: trypsin homology <TRY>
 F;18,208/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;21-151/Disulfide bonds: #status predicted
 F;25-158/Disulfide bonds: #status predicted
 F;70-86/Disulfide bonds: #status predicted
 F;85,139,238/Active site: His, Asp, Ser #status predicted
 F;175-244/Disulfide bonds: #status predicted
 F;207-223/Disulfide bonds: #status predicted
 F;234-264/Disulfide bonds: #status predicted

Query Match 21.5%; Score 503; DB 2; Length 418;
 Best Local Similarity 39.1%; Pred. No. 4.6e-34;
 Matches 108; Conservative 36; Mismatches 90; Indels 42; Gaps 7;
 QY 179 GPCSLGSLVSLHCLACGKSLKT-----GLTRQNSQAGTRIVGSGSAHVGAWPMWVSLQIFTSNRRY 67
 DB 23 GPC-----
 QY 226 HVCGSILDPHWLTAHCFRKHDTDFNWK-----VRAGSKLGSPF--SLAVAKIII 276
 DB 68 HACGGLLSHVLTAHCFDNKKKYDWRVLVFGAQEIEYGRNKPKEPQBERYVQKIVI 127
 QY 277 IEFNPMYKNDIALMKLOPPLTFSGTVRPICLPFDDELTATP--LWILGWGFTKQNG 334
 DB 128 HEKYNVTEGNDIALKLVPPVTCGNFPGCPCLPHFKAGPPRIKIPHTCTVTVGWGVIKREA 187
 QY 335 GKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEGGVDTCCGDSGGPLMYQSD 394
 DB 188 PRPSVLMARVLDLIDLCSNTQWNGRVTSTNCAGYPEGKIDTCQDSGGPLMCRDN 247
 QY 395 ---QHVGVGVSVNGYCGGSPSTPGVTTKVSAVLNWI 427
 DB 248 ARQPFVVVGITTSWGVGCARAKRPGVYTATWDYLDWI 283

RESULT 11
 S18407
 A;Title: Rat sperm acrosin: cDNA sequence, derived primary structure and phylogenetic of
 A;Reference number: S18407; MUID:92031708; PMID:1932123
 A;Accession: S18407
 A;Molecule type: mRNA
 A;Residues: 1-437 <KLE>
 A;Cross-references: EMBL:X59254
 R;Kleim, U.; Flake, A.; Engel, W.
 submitted to the EMBL Data Library, April 1991
 A;Reference number: S30037
 A;Accession: S30037
 A;Molecule type: mRNA
 A;Residues: 1-254, 'LCDR', '259', 'DHEL', '264', 'GRLC', '269-437' <KLE2>
 A;Cross-references: EMBL:X59254; NID:g57282; PIDN:CAA41947.1; PID:g57283
 R;Kremling, H.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.
 DNA Seq. 2, 57-60, 1991
 A;Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.
 A;Reference number: A56620; MUID:92199245; PMID:1802037
 A;Accession: A56620
 A;Status: preliminary
 A;Molecule type: DNA; mRNA
 A;Residues: 1-254, 'LCDR', '259', 'DHEL', '264', 'GRLC', '269-437' <KRE>
 A;Note: sequence modified after extraction from NCBI backbone
 A;Note: sequence extracted from NCBI backbone (NCBIN:89439, NCBIN:89447, NC
 C;Superfamily: acrosin; trypsin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen

F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-43/Domain: acrosin light chain #status predicted <LCH>
 F;43-286/Domain: trypsin homology <TRY>
 F;44-437/Domain: acrosin heavy chain #status predicted <HCH>
 F;22,211/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;25-155,29-163/Disulfide bonds: #status predicted
 F;74-90/Disulfide bonds: #status predicted
 F;89,143,241/Active site: His, Asp, Ser #status predicted
 F;178-247/Disulfide bonds: #status predicted
 F;210-226/Disulfide bonds: #status predicted
 F;237-267/Disulfide bonds: #status predicted

Query Match 21.4%; Score 501.5; DB 2; Length 437;
 Best Local Similarity 42.7%; Pred. No. 6.5e-34;
 Matches 105; Conservative 35; Mismatches 85; Indels 21; Gaps 6;
 QY 202 RVVGGEASVDSWPMQVSIQY-----DKQHVCGSILDPHWLTAHCFRKHDTDFNWK 255
 DB 42 RIVGGOTSSRAWPMWVSLQIFTSNRRYHACGGLLSHVLTAHCFDNKKKYDWR 101
 QY 256 -----VRAGSKLGSPFSLA--VAKIIIEFNPMYKNDIALMKLOPPLTFSGTVR 306
 DB 102 LVFGAHEIEYGRNKPKEPQBERYVQKIVHEKYNVTEGNDIALKLVPPVTCGDFVGP 161
 QY 307 ICLPFDDELTATP--LWILGWGFTKQNGKMSDILLOASVQVIDSTRCNADDAVQGEV 364
 DB 162 GCLPHF-KSGPPRIKIPHTCTVTVGWGVIKDNAPRSPVLMARVLDLIDLCSNTQWNGRV 220
 QY 365 TEKMKCAGIPEGGVDTCCGDSGGPLMYQSDQ---WHVGVGVSVNGYCGGSPSTPGVTTKV 421
 DB 221 TSTNVACGYPEGKIDTCQDSGGPLMCRDTRRPFVIVGITSWGVGCARAKRPGVYTATW 280
 QY 422 AYLNWI 427
 DB 281 DYLDWI 286

RESULT 12
 S29599
 A;Title: precursor - guinea pig (fragment)
 C;Species: Cavia porcellus (guinea pig)
 C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
 C;Accession: S29599
 R;Gerton, G.L.; Hoff, H.B.; Baba, T.
 submitted to the EMBL Data Library, May 1992
 A;Description: The amino acid sequence of guinea pig proacrosin deduced from its cDNA seq
 A;Reference number: S29599
 A;Accession: S29599
 A;Molecule type: mRNA
 A;Residues: 1-421 <GER>
 A;Cross-references: EMBL:Z12153; NID:g49559; PIDN:CAA78137.1; PID:g49560
 C;Superfamily: acrosin; trypsin homology
 C;Keywords: glycoprotein; hydrolase; serine proteinase
 F;41-284/Domain: trypsin homology <TRY>

Query Match 21.1%; Score 493.5; DB 2; Length 421;
 Best Local Similarity 38.0%; Pred. No. 2.9e-33;
 Matches 105; Conservative 44; Mismatches 84; Indels 43; Gaps 8;
 QY 179 GPC-----LSGSLVSLHCLACGKSLKTPRVVGGEASVDSWPMQVSIQY-----DKQ 225
 DB 25 GPCGLRFRQLQSGV-----RIIGGQTACQPGAWPMWVSLQIFMAHNNRY 69
 QY 226 HVCGSILDPHWLTAHCFRKHDTDFNWKVRAGSKL---GSFSLA-----VAKIII 276
 DB 70 HACGGLLSHVLTAHCFDNKKKYDWRVLVFGAQEIEYGRNKPKEPQBERYVQKIVI 129
 QY 277 IEFNPMYKNDIALMKLOPPLTFSGTVRPICLPFDDELTATP--LWILGWGFTKQNG 334
 DB 130 HEKYNVTEGNDIALKLVPPVTCGNFPGCPCLPHFKAGPPRIKIPHTCTVTVGWGVIKREA 188
 QY 335 GKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEGGVDTCCGDSGGPLMYQ-- 392

Db 189 PRPSVLLLEARVELIDLCLNCSQWYNGKVMSTNVCAGYPEGKIDTCQSGGGLMCRDN 248

QY 393 -SDQHHVGVISWVGCGGSPSTPGVYTKVSAVLNWI 427

Db 249 ANSPFVVGITSGVGGCARAKRPGIYTATWDYLDWI 284

RESULT 13

KOHUP

Plasma kallikrein (EC 3.4.21.34) precursor - human

N/Alternate names: Kiningenin; plasma prekallikrein

C/Species: Homo sapiens (man)

C/Date: 13-Aug-1986 #sequence revision 13-Aug-1986 #text_change 18-Jun-1999

C/Accession: A00921; A37939

R/Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

Biochemistry 25, 2410-2417, 1986

A/Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four b

A/Reference number: A00921; MUID:86243359; PMID:3521732

A/Accession: A00921

A/Molecule type: mRNA

A/Residues: 1-638 <CHU>

A/Cross-references: GB:M13143; NID:q190262; PIDN:AAA60153.1; PID:g190263

R/McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2050-2056, 1991

A/Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of

A/Reference number: A37939; MUID:91152016; PMID:1998666

A/Accession: A37939

A/Molecule type: protein

A/Residues: 20-27;40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80;103-113;131-140;141-142;143-144;145-146;147-148;149-150;151-152;153-154;155-156;157-158;159-160;161-162;163-164;165-166;167-168;169-170;171-172;173-174;175-176;177-178;179-180;181-182;183-184;185-186;187-188;189-190;191-192;193-194;195-196;197-198;199-200;201-202;203-204;205-206;207-208;209-210;211-212;213-214;215-216;217-218;219-220;221-222;223-224;225-226;227-228;229-230;231-232;233-234;235-236;237-238;239-240;241-242;243-244;245-246;247-248;249-250;251-252;253-254;255-256;257-258;259-260;261-262;263-264;265-266;267-268;269-270;271-272;273-274;275-276;277-278;279-280;281-282;283-284;285-286;287-288;289-290;291-292;293-294;295-296;297-298;299-300;301-302;303-304;305-306;307-308;309-310;311-312;313-314;315-316;317-318;319-320;321-322;323-324;325-326;327-328;329-330;331-332;333-334;335-336;337-338;339-340;341-342;343-344;345-346;347-348;349-350;351-352;353-354;355-356;357-358;359-360;361-362;363-364;365-366;367-368;369-370;371-372;373-374;375-376;377-378;379-380;381-382;383-384;385-386;387-388;389-390;391-392;393-394;395-396;397-398;399-400;401-402;403-404;405-406;407-408;409-410;411-412;413-414;415-416;417-418;419-420;421-422;423-424;425-426;427-428;429-430;431-432;433-434;435-436;437-438;439-440;441-442;443-444;445-446;447-448;449-450;451-452;453-454;455-456;457-458;459-460;461-462;463-464;465-466;467-468;469-470;471-472;473-474;475-476;477-478;479-480;481-482;483-484;485-486;487-488;489-490;491-492;493-494;495-496;497-498;499-500;501-502;503-504;505-506;507-508;509-510;511-512;513-514;515-516;517-518;519-520;521-522;523-524;525-526;527-528;529-530;531-532;533-534;535-536;537-538;539-540;541-542;543-544;545-546;547-548;549-550;551-552;553-554;555-556;557-558;559-560;561-562;563-564;565-566;567-568;569-570;571-572;573-574;575-576;577-578;579-580;581-582;583-584;585-586;587-588;589-590;591-592;593-594;595-596;597-598;599-600;601-602;603-604;605-606;607-608;609-610;611-612;613-614;615-616;617-618;619-620;621-622;623-624;625-626;627-628;629-630;631-632;633-634;635-636;637-638;639-640;641-642;643-644;645-646;647-648;649-650;651-652;653-654;655-656;657-658;659-660;661-662;663-664;665-666;667-668;669-670;671-672;673-674;675-676;677-678;679-680;681-682;683-684;685-686;687-688;689-690;691-692;693-694;695-696;697-698;699-700;701-702;703-704;705-706;707-708;709-710;711-712;713-714;715-716;717-718;719-720;721-722;723-724;725-726;727-728;729-730;731-732;733-734;735-736;737-738;739-740;741-742;743-744;745-746;747-748;749-750;751-752;753-754;755-756;757-758;759-760;761-762;763-764;765-766;767-768;769-770;771-772;773-774;775-776;777-778;779-780;781-782;783-784;785-786;787-788;789-790;791-792;793-794;795-796;797-798;799-800;801-802;803-804;805-806;807-808;809-810;811-812;813-814;815-816;817-818;819-820;821-822;823-824;825-826;827-828;829-830;831-832;833-834;835-836;837-838;839-840;841-842;843-844;845-846;847-848;849-850;851-852;853-854;855-856;857-858;859-860;861-862;863-864;865-866;867-868;869-870;871-872;873-874;875-876;877-878;879-880;881-882;883-884;885-886;887-888;889-890;891-892;893-894;895-896;897-898;899-900;901-902;903-904;905-906;907-908;909-910;911-912;913-914;915-916;917-918;919-920;921-922;923-924;925-926;927-928;929-930;931-932;933-934;935-936;937-938;939-940;941-942;943-944;945-946;947-948;949-950;951-952;953-954;955-956;957-958;959-960;961-962;963-964;965-966;967-968;969-970;971-972;973-974;975-976;977-978;979-980;981-982;983-984;985-986;987-988;989-990;991-992;993-994;995-996;997-998;999-1000;1001-1002;1003-1004;1005-1006;1007-1008;1009-1010;1011-1012;1013-1014;1015-1016;1017-1018;1019-1020;1021-1022;1023-1024;1025-1026;1027-1028;1029-1030;1031-1032;1033-1034;1035-1036;1037-1038;1039-1040;1041-1042;1043-1044;1045-1046;1047-1048;1049-1050;1051-1052;1053-1054;1055-1056;1057-1058;1059-1060;1061-1062;1063-1064;1065-1066;1067-1068;1069-1070;1071-1072;1073-1074;1075-1076;1077-1078;1079-1080;1081-1082;1083-1084;1085-1086;1087-1088;1089-1090;1091-1092;1093-1094;1095-1096;1097-1098;1099-1100;1101-1102;1103-1104;1105-1106;1107-1108;1109-1110;1111-1112;1113-1114;1115-1116;1117-1118;1119-1120;1121-1122;1123-1124;1125-1126;1127-1128;1129-1130;1131-1132;1133-1134;1135-1136;1137-1138;1139-1140;1141-1142;1143-1144;1145-1146;1147-1148;1149-1150;1151-1152;1153-1154;1155-1156;1157-1158;1159-1160;1161-1162;1163-1164;1165-1166;1167-1168;1169-1170;1171-1172;1173-1174;1175-1176;1177-1178;1179-1180;1181-1182;1183-1184;1185-1186;1187-1188;1189-1190;1191-1192;1193-1194;1195-1196;1197-1198;1199-1200;1201-1202;1203-1204;1205-1206;1207-1208;1209-1210;1211-1212;1213-1214;1215-1216;1217-1218;1219-1220;1221-1222;1223-1224;1225-1226;1227-1228;1229-1230;1231-1232;1233-1234;1235-1236;1237-1238;1239-1240;1241-1242;1243-1244;1245-1246;1247-1248;1249-1250;1251-1252;1253-1254;1255-1256;1257-1258;1259-1260;1261-1262;1263-1264;1265-1266;1267-1268;1269-1270;1271-1272;1273-1274;1275-1276;1277-1278;1279-1280;1281-1282;1283-1284;1285-1286;1287-1288;1289-1290;1291-1292;1293-1294;1295-1296;1297-1298;1299-1300;1301-1302;1303-1304;1305-1306;1307-1308;1309-1310;1311-1312;1313-1314;1315-1316;1317-1318;1319-1320;1321-1322;1323-1324;1325-1326;1327-1328;1329-1330;1331-1332;1333-1334;1335-1336;1337-1338;1339-1340;1341-1342;1343-1344;1345-1346;1347-1348;1349-1350;1351-1352;1353-1354;1355-1356;1357-1358;1359-1360;1361-1362;1363-1364;1365-1366;1367-1368;1369-1370;1371-1372;1373-1374;1375-1376;1377-1378;1379-1380;1381-1382;1383-1384;1385-1386;1387-1388;1389-1390;1391-1392;1393-1394;1395-1396;1397-1398;1399-1400;1401-1402;1403-1404;1405-1406;1407-1408;1409-1410;1411-1412;1413-1414;1415-1416;1417-1418;1419-1420;1421-1422;1423-1424;1425-1426;1427-1428;1429-1430;1431-1432;1433-1434;1435-1436;1437-1438;1439-1440;1441-1442;1443-1444;1445-1446;1447-1448;1449-1450;1451-1452;1453-1454;1455-1456;1457-1458;1459-1460;1461-1462;1463-1464;1465-1466;1467-1468;1469-1470;1471-1472;1473-1474;1475-1476;1477-1478;1479-1480;1481-1482;1483-1484;1485-1486;1487-1488;1489-1490;1491-1492;1493-1494;1495-1496;1497-1498;1499-1500;1501-1502;1503-1504;1505-1506;1507-1508;1509-1510;1511-1512;1513-1514;1515-1516;1517-1518;1519-1520;1521-1522;1523-1524;1525-1526;1527-1528;1529-1530;1531-1532;1533-1534;1535-1536;1537-1538;1539-1540;1541-1542;1543-1544;1545-1546;1547-1548;1549-1550;1551-1552;1553-1554;1555-1556;1557-1558;1559-1560;1561-1562;1563-1564;1565-1566;1567-1568;1569-1570;1571-1572;1573-1574;1575-1576;1577-1578;1579-1580;1581-1582;1583-1584;1585-1586;1587-1588;1589-1590;1591-1592;1593-1594;1595-1596;1597-1598;1599-1600;1601-1602;1603-1604;1605-1606;1607-1608;1609-1610;1611-1612;1613-1614;1615-1616;1617-1618;1619-1620;1621-1622;1623-1624;1625-1626;1627-1628;1629-1630;1631-1632;1633-1634;1635-1636;1637-1638;1639-1640;1641-1642;1643-1644;1645-1646;1647-1648;1649-1650;1651-1652;1653-1654;1655-1656;1657-1658;1659-1660;1661-1662;1663-1664;1665-1666;1667-1668;1669-1670;1671-1672;1673-1674;1675-1676;1677-1678;1679-1680;1681-1682;1683-1684;1685-1686;1687-1688;1689-1690;1691-1692;1693-1694;1695-1696;1697-1698;1699-1700;1701-1702;1703-1704;1705-1706;1707-1708;1709-1710;1711-1712;1713-1714;1715-1716;1717-1718;1719-1720;1721-1722;1723-1724;1725-1726;1727-1728;1729-1730;1731-1732;1733-1734;1735-1736;1737-1738;1739-1740;1741-1742;1743-1744;1745-1746;1747-1748;1749-1750;1751-1752;1753-1754;1755-1756;1757-1758;1759-1760;1761-1762;1763-1764;1765-1766;1767-1768;1769-1770;1771-1772;1773-1774;1775-1776;1777-1778;1779-1780;1781-1782;1783-1784;1785-1786;1787-1788;1789-1790;1791-1792;1793-1794;1795-1796;1797-1798;1799-1800;1801-1802;1803-1804;1805-1806;1807-1808;1809-1810;1811-1812;1813-1814;1815-1816;1817-1818;1819-1820;1821-1822;1823-1824;1825-1826;1827-1828;1829-1830;1831-1832;1833-1834;1835-1836;1837-1838;1839-1840;1841-1842;1843-1844;1845-1846;1847-1848;1849-1850;1851-1852;1853-1854;1855-1856;1857-1858;1859-1860;1861-1862;1863-1864;1865-1866;1867-1868;1869-1870;1871-1872;1873-1874;1875-1876;1877-1878;1879-1880;1881-1882;1883-1884;1885-1886;1887-1888;1889-1890;1891-1892;1893-1894;1895-1896;1897-1898;1899-1900;1901-1902;1903-1904;1905-1906;1907-1908;1909-1910;1911-1912;1913-1914;1915-1916;1917-1918;1919-1920;1921-1922;1923-1924;1925-1926;1927-1928;1929-1930;1931-1932;1933-1934;1935-1936;1937-1938;1939-1940;1941-1942;1943-1944;1945-1946;1947-1948;1949-1950;1951-1952;1953-1954;1955-1956;1957-1958;1959-1960;1961-1962;1963-1964;1965-1966;1967-1968;1969-1970;1971-1972;1973-1974;1975-1976;1977-1978;1979-1980;1981-1982;1983-1984;1985-1986;1987-1988;1989-1990;1991-1992;1993-1994;1995-1996;1997-1998;1999-2000;2001-2002;2003-2004;2005-2006;2007-2008;2009-2010;2011-2012;2013-2014;2015-2016;2017-2018;2019-2020;2021-2022;2023-2024;2025-2026;2027-2028;2029-2030;2031-2032;2033-2034;2035-2036;2037-2038;2039-2040;2041-2042;2043-2044;2045-2046;2047-2048;2049-2050;2051-2052;2053-2054;2055-2056;2057-2058;2059-2060;2061-2062;2063-2064;2065-2066;2067-2068;2069-2070;2071-2072;2073-2074;2075-2076;2077-2078;2079-2080;2081-2082;2083-2084;2085-2086;2087-2088;2089-2090;2091-2092;2093-2094;2095-2096;2097-2098;2099-2100;2101-2102;2103-2104;2105-2106;2107-2108;2109-2110;2111-2112;2113-2114;2115-2116;2117-2118;2119-2120;2121-2122;2123-2124;2125-2126;2127-2128;2129-2130;2131-2132;2133-2134;2135-2136;2137-2138;2139-2140;2141-2142;2143-2144;2145-2146;2147-2148;2149-2150;2151-2152;2153-2154;2155-2156;2157-2158;2159-2160;2161-2162;2163-2164;2165-2166;2167-2168;2169-2170;2171-2172;2173-2174;2175-2176;2177-2178;2179-2180;2181-2182;2183-2184;2185-2186;2187-2188;2189-2190;2191-2192;2193-2194;2195-2196;2197-2198;2199-2200;2201-2202;2203-2204;2205-2206;2207-2208;2209-2210;2211-2212;2213-2214;2215-2216;2217-2218;2219-2220;2221-2222;2223-2224;2225-2226;2227-2228;2229-2230;2231-2232;2233-2234;2235-2236;2237-2238;2239-2240;2241-2242;2243-2244;2245-2246;2247-2248;2249-2250;2251-2252;2253-2254;2255-2256;2257-2258;2259-2260;2261-2262;2263-2264;2265-2266;2267-2268;2269-2270;2271-2272;2273-2274;2275-2276;2277-2278;2279-2280;2281-2282;2283-2284;2285-2286;2287-2288;2289-2290;2291-2292;2293-2294;2295-2296;2297-2298;2299-2300;2301-2302;2303-2304;2305-2306;2307-2308;2309-2310;2311-2312;2313-2314;2315-2316;2317-2318;2319-2320;2321-2322;2323-2324;2325-2326;2327-2328;2329-2330;2331-2332;2333-2334;2335-2336;2337-2338;2339-2340;2341-2342;2343-2344;2345-2346;2347-2348;2349-2350;2351-2352;2353-2354;2355-2356;2357-2358;2359-2360;2361-2362;2363-2364;2365-2366;2367-2368;2369-2370;2371-2372;2373-2374;2375-2376;2377-2378;2379-2380;2381-2382;2383-2384;2385-2386;2387-2388;2389-2390;2391-2392;2393-2394;2395-2396;2397-2398;2399-2400;2401-2402;2403-2404;2405-2406;2407-2408;2409-2410;2411-2412;2413-2414;2415-2416;2417-2418;2419-2420;2421-2422;2423-2424;2425-2426;2427-2428;2429-2430;2431-2432;2433-2434;2435-2436;2437-2438;2439-2440;2441-2442;2443-2444;2445-2446;2447-2448;2449-2450;2451-2452;2453-2454;2455-2456;2457-2458;2459-2460;2461-2462;2463-2464;2465-2466;2467-2468;2469-2470;2471-2472;2473-2474;2475-2476;2477-2478;2479-2480;2481-2482;2483-2484;2485-2486;2487-2488;2489-2490;2491-2492;2493-2494;2495-2496;2497-2498;2499-2500;2501-2502;2503-2504;2505-2506;2507-2508;2509-2510;2511-2512;2513-2514;2515-2516;2517-2518;2519-2520;2521-2522;2523-2524;2525-2526;2527-2528;2529-2530;2531-2532;2533-2534;2535-2536;2537-2538;2539-2540;2541-2542;2543-2544;2545-2546;2547-2548;2549-2550;2551-2552;2553-2554;2555-2556;2557-2558;2559-2560;2561-2562;2563-2564;2565-2566;2567-2568;2569-2570;2571-2572;2573-2574;2575-2576;2577-2578;2579-2580;2581-2582;2583-2584;2585-2586;2587-2588;2589-2590;2591-2592;2593-2594;2595-2596;2597-2598;2599-2600;2601-2602;2603-2604;2605-2606;2607-2608;2609-2610;2611-2612;2613-2614;2615-2616;2617-2618;2619-2620;2621-2622;2623-

A:Accession: S11674
A:Molecule type: DNA
A:Residues: 1-421 <KEI>
A:Cross-references: EMBL:X54017; NID:G35592; PIDN:CAA37964.1; PID:G1216165
A:Note: the authors translated the codon AGG for residue 64 as Thr and CTG for residue 2
R;Vazquez-Levin, M.H.; Reventosa, J.; Gordon, J.W.
Eur. J. Biochem. 207, 23-26, 1992
A:Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence
A:Reference number: S23499; MUID:92331659; PMID:1628652
A:Accession: S23499
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-421 <VAZ>
A:Cross-references: EMBL:M77378
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
R;Keime, S.
submitted to the EMBL Data Library, December 1989
A:Reference number: S12063
A:Accession: S12063
A:Molecule type: DNA
A:Residues: 1-225, 'R', 227-421 <KEI2>
A:Cross-references: EMBL:X54017
R;Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
Hum. Genet. 84, 125-128, 1990
A:Title: Molecular cloning of human proacrosin cDNA.
A:Reference number: A61022; MUID:90128988; PMID:2298447
A:Accession: A61022
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
R;Raba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
FEBS Lett. 244, 296-300, 1989
A:Title: Primary structure of human proacrosin deduced from its cDNA sequence.
A:Reference number: S03330; MUID:89153568; PMID:2493394
A:Accession: S03330
A:Molecule type: mRNA
A:Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
A:Cross-references: EMBL:Y00970; NID:G28325; PIDN:CAA68784.1; PID:G28326
C:Genetics:
A:Gene: GDB:ACR
A:Cross-references: GDB:119645; OMIM:102480
A:Map position: 22q13-22qter
A:Introns: 26/2; 94/2; 189/2; 237/3
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-421/Product: acrosin #status predicted <MAT>
F;20-42/Product: acrosin light chain #status predicted <LCH>
F;43-421/Product: acrosin heavy chain #status predicted <HCH>
F;43-285/Domain: trypsin homology <TRY>
F;102-379/Region: proline-rich
F;22-210/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;25-154/Disulfide bonds: #status predicted
F;29-162/Disulfide bonds: #status predicted
F;73-89/Disulfide bonds: #status predicted
F;88-142,240/Active site: His, Asp, Ser #status predicted
F;177-246/Disulfide bonds: #status predicted
F;209-225/Disulfide bonds: #status predicted
F;236-266/Disulfide bonds: #status predicted

Query Match 20.5%; Score 481; DB 1; Length 421;
Best Local Similarity 41.0%; Pred. No. 3.1e-32;
Matches 100; Conservative 38; Mismatches 88; Indels 18; Gaps 5;

Qy 202 RVVGGEEASVDSWPVOVSIQY-----DKOHVGGSLDPHWLTAHCFKHTDVFNNKV 256
Db 42 RVVGGAAQHGAWPWSVQIFRYNSHRYHTCGSLNSRWLTAHCFVGNVHDWRL 101
Qy 257 RAGSDXL---GSPPSLA-----VAKIIIEFNPMYFKNDIALMKLPFLFSGTVPRI 307
Db 102 VFGAKBITYGNKPKVAPLQERYVEKIIIEHKYNSATEGNDIALVEITPISCGRFPG 161
Qy 308 CLPFFDEELTPAT-PLWITGWFTKQNGGKWDILLQASVOVIDSTRCNADDAYQGEVTE 366

162 CLPHFKAGLPRGSGSCWVAGWGYIEEKAPRPSSILMEARVDLIDLDCNSTOWNGRVOP 221
367 KMCAGIPEGGVDTCCGDSGGPLMYQSDQ---WHVVGIVSWGYGCGSPSTPGVYTKVSAY 423
222 TNVCAGYEVGKIDTCGDSGGPLMKCKDSKESAYVVVGITISWGVGCALAKRPGIYATWPY 281
424 LNWI 427
282 LNWI 285

Search completed: June 7, 2004, 08:41:05
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 18 Seconds
(without alignments)
1258.362 Million cell updates/sec

Title: US-10-030-688-2
Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVPLKRP.....VYTKVSAYLNWYNVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2337	99.8	437	1 TMS4 HUMAN	Q9nr54 homo sapien
2	1844	78.7	435	1 TMS4 MOUSE	O8vca5 mus musculus
3	887	28.3	452	1 TMS3 MOUSE	O8k1t0 mus musculus
4	673.5	28.8	493	1 TMS2 HUMAN	O15393 homo sapien
5	660.5	28.2	454	1 TMS3 HUMAN	P57727 homo sapien
6	648.5	27.7	490	1 TMS2 MOUSE	Q91jq8 mus musculus
7	628.5	26.8	455	1 TMS5 MOUSE	O9er04 mus musculus
8	584.5	25.0	417	1 HEP5 HUMAN	P05981 homo sapien
9	582.5	24.9	1019	1 ENTK HUMAN	P98073 homo sapien
10	578	24.7	436	1 HEP5 MOUSE	O35453 mus musculus
11	574	24.5	457	1 TMS5 HUMAN	Q9h3s3 homo sapien
12	571	24.4	1035	1 ENTK BOVIN	P98072 bos taurus
13	568.5	24.3	416	1 HEP5 RAT	Q05511 rattus norv
14	568	24.3	418	1 HATT HUMAN	O60235 homo sapien
15	552.5	23.6	1069	1 ENTK MOUSE	P97435 mus musculus
16	552	23.6	1034	1 ENTK PIG	P98074 sus scrofa
17	550.5	23.5	811	1 TMS6 MOUSE	Q9dbi0 mus musculus
18	531	22.7	811	1 TMS6 HUMAN	Q8tu80 homo sapien
19	511	21.8	638	1 KAL RAT	P14272 rattus norv
20	508	21.7	638	1 KAL MOUSE	P26262 mus musculus
21	505	21.6	436	1 ACRO MOUSE	P23157 mus musculus
22	503.5	21.5	324	1 TEST MOUSE	Q91bj7 mus musculus
23	501.5	21.4	437	1 ACRO RAT	P22933 rattus norv
24	498	21.3	875	1 NETR HUMAN	P03952 homo sapien
25	491.5	21.0	638	1 KAL HUMAN	Q92319 mus musculus
26	488.5	20.9	1113	1 CORI MOUSE	Q9y5q5 homo sapien
27	484	20.7	1042	1 CORI HUMAN	Q9nr12 homo sapien
28	483	20.6	321	1 TRYG HUMAN	Q91bj7 mus musculus
29	482	20.6	855	1 ST14 MOUSE	P56677 mus musculus
30	481	20.5	421	1 ACRO HUMAN	P10323 homo sapien
31	478.5	20.4	415	1 ACRO PIG	P08001 sus scrofa
32	478	20.4	314	1 TEST HUMAN	Q9y6m0 homo sapien
33	476.5	20.3	311	1 TRYG MOUSE	Q9qu17 mus musculus

34	475	20.3	422	1 DES1 HUMAN	O9ul52 homo sapien
35	472.5	20.2	317	1 BSS4 HUMAN	O9gn44 homo sapien
36	470	20.1	761	1 NETR MOUSE	O08762 mus musculus
37	468.5	20.0	431	1 ACRO RABIT	P48038 oryctolagus
38	467.5	20.0	290	1 PR27 HUMAN	O9bgr3 homo sapien
39	463	19.8	855	1 ST14 HUMAN	O9y5v6 homo sapien
40	461.5	19.7	277	1 TRY2 ANOGA	P35036 anopheles g
41	460.5	19.7	263	1 CTRB RAT	P07338 rattus norv
42	459.5	19.6	342	1 PSS8 RAT	O9e897 rattus norv
43	459.5	19.6	812	1 PLMN MOUSE	P20918 mus musculus
44	458.5	19.6	267	1 TRY7 ANOGA	P35041 anopheles g
45	458.5	19.6	790	1 PLMN PIG	P06867 sus scrofa

ALIGNMENTS

RESULT 1
TMS4 HUMAN
ID TMS4 HUMAN STANDARD; PRT; 437 AA.
AC Q9NR54; Q9NZAS;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Membrane-type serine
protease 2) (MT-SP2).
GN TMSRS4 OR TMSRS3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OK NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Pancratic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Wallrapp C., Hasbnel S., Mueller-Pillasch F., Burghardt B.,
RA Iwamura T., Ruthenburger M., Lerch M.M., Adler G., Gress T.M.;
RT "A novel transmembrane serine protease (TMSRS3) overexpressed in
pancreatic cancer";
RL Cancer Res. 60:2602-2606(2000).
[2]
SEQUENCE FROM N.A.
RP Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
trachea, colon, and small intestine: identification, cloning, and
chromosomal localization.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
ENaC (By similarity).

CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: High levels in pancreatic, gastric, colorectal
 CC and ampullary cancer. Very weak expression in normal
 CC gastrointestinal and urogenital tract.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: AF179224; AAF74526.1; -;
 CC EMBL: AF216312; AAF31436.1; -;
 CC EMBL: BC011703; AAF11703.1; -;
 CC HSP: P00763; IDPO.
 CC MEROPS: S01.034; -;
 CC Genew: HGNC:11878; TMPSR54.
 CC MIM: 605563; -;
 CC GO: GO:0016021; C: integral to membrane; NAS.
 CC GO: GO:0004252; P: serine-type endopeptidase activity; NAS.
 CC GO: GO:0006508; P: proteolysis and peptidolysis; NAS.
 CC InterPro: IPR009003; Cys Ser trypsin.
 CC InterPro: IPR002172; LDL receptor A.
 CC InterPro: IPR001254; Peptidase S1.
 CC InterPro: IPR001344; Peptidase SLA.
 CC InterPro: IPR001190; Srcr receptor.
 CC Pfam: PF00057; ldl_recept_a; 1.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC SMART: SM00192; LDLa; 1.
 CC SMART: SM00202; SR; 1.
 CC SMART: SM00202; Tryp_Spc; 1.
 CC PROSITE: PS01209; LDLRA_1; FALSE NEG.
 CC PROSITE: PS00668; LDLRA_2; FALSE NEG.
 CC PROSITE: PS00420; SRCR_1; FALSE NEG.
 CC PROSITE: PS0287; SRCR_2; 1.
 CC PROSITE: PS0240; TRYPSIN_DOM; 1.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 KW DOMAIN 1 32 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT TRANSMEM 33 53 (POTENTIAL).
 FT
 FT DOMAIN 54 437 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 61 93 LDL-RECEPTOR CLASS A.
 FT DOMAIN 94 204 SRCR.
 FT DOMAIN 205 437 SERINE PROTEASE.
 FT ACT_SITE 245 245 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 204 205 CLEAVAGE (POTENTIAL).
 FT DISULFID 64 83 BY SIMILARITY.
 FT DISULFID 77 92 BY SIMILARITY.
 FT DISULFID 127 183 BY SIMILARITY.
 FT DISULFID 140 193 BY SIMILARITY.
 FT DISULFID 196 310 BY SIMILARITY.
 FT DISULFID 230 246 BY SIMILARITY.
 FT DISULFID 356 372 BY SIMILARITY.
 FT DISULFID 383 410 BY SIMILARITY.
 FT CARBOHYD 130 130 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 1 31 MFLQPSDQPLNSLDVPLKRPDMETFRK -> MSNPCA
 FT
 FT SEQUENCE 437 AA; 48204 MW; 351B2FD4A8657B12 CRC64;
 Query Match 99.8%; Score 2137; DB 1; Length 437;
 Best Local Similarity 100.0%; Pred. No. 4.2e-194; Indels 0; Gaps 0;
 Matches 434; Conservative 0; Mismatches 0;

QY 2 DPDSQPLNSLDVPLKRPDMETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 61
 DB 4 DPDSQPLNSLDVPLKRPDMETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 63
 QY 62 CGOPLHFIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 121
 DB 64 CGOPLHFIPRKQLCDGELDCPLGEDEEHCVKSPPEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFDNFTALAEACRQMGYSKPTFRFVEGPDDLDVVEITNSQELNRNSSGPC 181
 DB 124 FSACFDNFTALAEACRQMGYSKPTFRFVEGPDDLDVVEITNSQELNRNSSGPC 183
 QY 182 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPQVSIQYDKQHVCGGSLDPHVLTA 241
 DB 184 LSGSLVSLHCLACGKSLKTRVVGGEASVDSWPQVSIQYDKQHVCGGSLDPHVLTA 243
 QY 242 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFPMYKNDIALMKLOFLTPS 301
 DB 244 AHCFRKHDTVFNKVRAGSKLGSFSLAVAKIIIEFPMYKNDIALMKLOFLTPS 303
 QY 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 361
 DB 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKONGKMSDILLQASVQVIDSTRCNADDAVQ 363
 QY 362 GEVTEKNCAGIPEGGVDTCCQDSGGLMYQSDQHWVGVISWGVGCGSPSPGVTKVS 421
 DB 364 GEVTEKNCAGIPEGGVDTCCQDSGGLMYQSDQHWVGVISWGVGCGSPSPGVTKVS 423
 QY 422 AYLNWIYVWKAEI 435
 DB 424 AYLNWIYVWKAEI 437
 RESULT 2
 TMS4 MOUSE
 ID TMS4 MOUSE STANDARD; PRT; 435 AA.
 AC Q8VCA5; 2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transmembrane protease, serine 4 (EC 3.4.21.-) (Channel-activating
 DE protease 2) (mCAP2).
 GN TMPSR54 OR CAP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RP MEDLINE=22144321; PubMed=12149280;
 RA Vuagniaux G., Vallet V., Jaeger N.F., Hummler E., Rossier B.C.;
 RT "Synergistic activation of ENAC by three membrane-bound channel-
 RT activating serine proteases (mCAP1, mCAP2, and mCAP3) and serum- and
 RT glucocorticoid-regulated kinase (Sgk1) in Xenopus oocytes.";
 RL J. Gen. Physiol. 120:191-201(2002).
 RN (2)
 RP SEQUENCE FROM N.A.
 RP TISSUE=Breast tumor;
 RC MEDLINE=22388257; PubMed=12477932;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
CC ENAC.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AY043240; AK85307.1; -;
CC EMBL; BC021368; AAH21368.1; -;
CC HSSP; P00761; IAN1.
CC -----
CC MGD; MGI:2384877; Tmprss4.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR001190; Srcr receptor.
CC Pfam; PF00057; ldl_recept_a; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PROSITE; PS01209; LDLRA_1; FALSE NEG.
CC PROSITE; PS0068; LDLRA_2; FALSE NEG.
CC PROSITE; PS00420; SRCR_1; FALSE NEG.
CC PROSITE; PS0287; SRCR_2; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT DOMAIN 1 30 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 31 51 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT -----
FT DOMAIN 52 435 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 59 101 LDL-RECEPTOR CLASS A.
FT DOMAIN 102 202 SRCR
FT DOMAIN 203 435 SERINE PROTEASE.
FT ACT_SITE 243 243 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 288 288 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 385 385 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 202 203 CLEAVAGE (POTENTIAL).
FT DISULFID 62 81 BY SIMILARITY.
FT DISULFID 75 90 BY SIMILARITY.
FT DISULFID 125 181 BY SIMILARITY.
FT DISULFID 138 191 BY SIMILARITY.
FT DISULFID 194 308 BY SIMILARITY.
FT DISULFID 228 244 BY SIMILARITY.
FT DISULFID 354 370 BY SIMILARITY.
FT DISULFID 381 408 BY SIMILARITY.
FT CARBOHYD 128 128 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 435 AA; 47495 MW; DC52E45A43E01369 CRC64;
Query Match 78.7%; Score 1844; DB 1; Length 435;
Best Local Similarity 77.2%; Pred. No. 1.5e-151;
Matches 336; Conservative 43; Mismatches 56; Indels 0; Gaps 0;
QY 1 MDPDSQPLNSLDVKLRPRIPMETFRKGVPIIALLSLASIIIVVLIKVLDKYVF 60
DB 1 MESDSQGPLNRRDIVFRKPRPQETFKVGIPITIAVLLSLIAVALLIKVLDKYVF 60

QY 61 LCQGPLHFIPRKOLCDGELDCPLGEDEHCVKSFPEGPAAVAVRLSKDRSTLQVLDATGN 120
DB 61 ICSPPTFFIORGLCDGHLDCASGEDEHCVKDFPEKPGVAVRLSKDRSTLQVLDATGT 120
QY 121 WFSACFDNFTEALAEACRQMGYSKPTFRVIGIPDQDLVDVVEITENSQELRMNRSSGP 180
DB 121 WASVCFDNFTEALAKTACRQMGYSQPAFRAVEIRPDQLPVAQVTGNSQELQVNGRSR 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSIIDPHWILT 240
DB 181 CLSGSLVSLRCLDCGSLKTPRVVGGVEAPVDSWPQVSIQYKQHVCGGSIIDPHWILT 240
QY 241 AAHCFRKHTDVFVWKVRAGSDKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPLTF 300
DB 241 AAHCFRKYLDSWKVRAGSNILGNSPSLPVAKIFIAEPNPLYPKEKDIALVKLOPLTF 300
QY 301 SGTVRPICLPFBELTPATPLWILGFTFGKNGKMSDILLGASVQVVDSTECNADDA 360
DB 301 SGSVRPCLPFBELVLPATPVWVIGFTFGKNGKMSDILLGASVQVVDSTECNADDA 360
QY 361 QGEVTEKMCAGIPGEGVDTCCQDSDGGLPMYQSDQHVHVGIVSGVGCSPGPGVYTKV 420
DB 361 EGEVTAEMLCAGTFQGGKDTCCQDSDGGLPMYHSDKQVGVIVSGHCGCGSPGPGVYTKV 420
QY 421 SAYLNWIYNYWKAEL 435
DB 421 TAYLNWIYNYRKSEM 435
RESULT 3
TMS3 MOUSE STANDARD; PRT; 453 AA.
ID TMS3 MOUSE Q8K1T0; Q8VDE0;
AC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 3 (EC 3.4.21.-).
GN TMSRSS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
RX MEDLINE=22281255; PubMed=12393794;
RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
RA Buchet K., Raymond A., Hummler E., Marzella P.L., Kudoh J.,
RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
RT "The transmembrane serine protease (TMSRSS3) mutated in deafness
DFNB8/10 activates the epithelial sodium channel (ENAC) in vitro";
RL Hum. Mol. Genet. 11:2829-2836 (2002).
CC -!- FUNCTION: Probable protease. Seems to be capable of activating
CC ENAC.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC -!- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
CC supporting the organ of Corti and the stria vascularis.
CC -!- PTM: Undergoes autotryptic activation.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
DR EMBL; AJ429216; CAD22137.1; -
DR EMBL; AJ300738; CAC83350.1; -
DR HSP; P00761; IAN1.
DR MGD; MG1:2155445; Tmpss3.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; P00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; TYP SPC; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00688; LDLRA_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS0240; TRYPSIN DOM; 1.
DR PROSITE; PS00134; TRYPSIN HIS; 1.
DR PROSITE; PS00135; TRYPSIN SR; 1.
DR Hydrolase; Serine protease; Signal-anchor; Zymogen;
KW Endoplasmic reticulum.
FT DOMAIN 1 48
FT TRANSMEM 49 69
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT POTENTIAL.
FT FT 70 453
FT DOMAIN 72 108
FT DOMAIN 104 205
FT DOMAIN 217 448
FT ACT_SITE 257 257
FT ACT_SITE 304 304
FT ACT_SITE 400 400
FT SITE 216 217
FT DISULFID 73 85
FT DISULFID 79 98
FT DISULFID 92 107
FT DISULFID 129 184
FT DISULFID 142 204
FT DISULFID 207 324
FT DISULFID 242 258
FT DISULFID 338 406
FT DISULFID 369 385
FT DISULFID 396 424
FT CARBOHYD 221 221
FT CONFLICT 117 117
FT CONFLICT 246 246
FT SEQUENCE 453 AA; 49491 MW; 1ABCBF0AF6E1EF6 CRC64;
Query Match 29.3%; Score 687; DB 1; Length 453;
Best Local Similarity 36.9%; Pred. No. 1.2e-51;
Matches 161; Conservative 72; Mismatches 153; Indels 50; Gaps 14;
QY 22 IPMETFRKVGIPITIIALLSLASIIIVVLKIVLD---KYFLCQOPLHIFPRKQLCDGE 78
DB 44 LFLKFFPIIVIGIALLALA-----ICLGHFDCSGKY--RCHSFKFELTARCDGV 95
QY 79 LDCPLGEDEHCKVFPPEGPAVAVLSKDRSTLQVLDATGNWSACPDNTEALAEAC 138
DB 96 SCKNAEDEYRC-----VRVSGRAALQVFTAAA--WRFTCSDDDKSHYAKIAC 142
QY 139 ROMGYSS-----KPTFAVEIGCPDQDLDWZETITENSQELRNRSSGPCLSGL 186
DB 143 AQLGFPYSVSSDLHRLVDALEBQFGDFVSINHLSDRVTALHSHVYNRG--CTSGHV 199
QY 187 VSLHCLAGC-KSLKTPRVVGGEEASVDSWPQVSTQYDKQVCGGSLDHPWVTAHCF 245
DB 200 VTLKCSAGCTRTGYSGRIVGNGMSSLQWPMQVSLQFGYHLGCGSVITPLWIVTAHCV 259
QY 246 RKHTDVFNNKVRAGSKL--GSFPLSLAVAKIIITFENPMY-PK--DNDIALMKLQFP 300
DB 260 YDLHYPKSNVTQVGLVSLMDSVPVPSHLEVKII---YHSKYFKRLGNDIALMKLSEPLTP 316
```

```
QY 301 SGTVRPICLPFFDELTPTATPLWIGWTKQNGKMSDILLOASVOVIDSTRCNADDAY 360
DB 317 DETIOPICLPNSENSEFPDGKLCWTSGWAT-EDGDRASPVLNHAAPLISNKICNHRD 375
QY 361 QGEVTERKMCAGIPEGGVDTCQDGGGLMYQSDQ-WHVVGVISWGYCGGPGSTPGYTK 419
DB 376 GGIISPSMLCAGYLKGGVDSQDGGGLVCQERRLWLKVGATSGFICGAENVKPGVYTR 435
QY 420 VSAYLWYINWKAEL 435
DB 436 ITSFLDWIHEQLERDL 451
RESULT 4
TMS2_HUMAN
ID TMS2_HUMAN STANDARD; PRT; 492 AA.
AC OLS393; Q9BX1;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane protease, serine 2 precursor (EC 3.4.21.-).
GN TMRSS2 OR PRSS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=97458144; PubMed=9325052;
RA Paoloni-Giacobino A., Chen H., Feitsch M.C., Rossier C.,
Antanarakis S.E.;
"Cloning of the TMRSS2 gene, which encodes a novel serine protease
with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";
Genomics 44:309-320(1997).
[2]
SEQUENCE FROM N.A.
MEDLINE=21309069; PubMed=11414763;
RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
"Mutation analyses of 268 candidate genes in human tumor cell lines.";
Genomics 74:352-364(2001).
[3]
SEQUENCE FROM N.A. AND MUTAGENESIS.
MEDLINE=21139112; PubMed=11245484;
RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,
Raitano A.B., Jakobovits A.;
"Catalytic cleavage of the androgen-regulated TMRSS2 protease results
in its secretion by prostate and prostate cancer epithelia.";
Cancer Res. 61:1686-1692(2001).
[4]
TISSUE SPECIFICITY
MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikho P.T.;
"Expression of transmembrane serine protease TMRSS2 in mouse and
human tissues.";
J. Pathol. 193:134-140(2001).
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
cleavage and secreted.
CC -I- TISSUE SPECIFICITY: Expressed strongly in small intestine. Also
expressed in prostate, colon, stomach and salivary gland.
CC -I- SIMILARITY: Belongs to peptidase family S1.
CC -I- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -I- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U75329; AAC51784.1; -
CC EMBL; AF123453; AAD31117.1; -
DR
```

DR EMBL; AF270487; RAK29280.1; --
 DR HSPB; P00763; IDPO.
 DR MEROPS; S01.247; --
 DR Genew; HGNC:11876; TMPRSS2.
 DR MIM; 602060; --
 DR GO; GO:0003887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008236; F:serine-type peptidase activity; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS00668; LDLRA_2; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS02987; SRCR_2; 1.
 DR PROSITE; PS02400; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Signal-anchor; Zymogen;
 KW Polymorphism.
 FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-
 FT CHAIN 256 492 CATALYTIC CHAIN.
 FT CHAIN 492 TRANSMEMBRANE PROTEASE, SERINE 2,
 FT CHAIN 492 CATALYTIC CHAIN.
 FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 112 149 LDL-RECEPTOR CLASS A.
 FT DOMAIN 150 242 SRCR.
 FT DOMAIN 256 492 SERINE PROTEASE.
 FT ACT_SITE 296 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 441 441 CHARGE RELAY SYSTEM.
 FT SITE 255 256 CLEAVAGE (POTENTIAL).
 FT DISULFID 113 126 BY SIMILARITY.
 FT DISULFID 120 139 BY SIMILARITY.
 FT DISULFID 133 148 BY SIMILARITY.
 FT DISULFID 172 231 BY SIMILARITY.
 FT DISULFID 185 241 BY SIMILARITY.
 FT DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 281 297 BY SIMILARITY.
 FT DISULFID 410 426 BY SIMILARITY.
 FT DISULFID 437 465 BY SIMILARITY.
 FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 449 449 K -> N (in dBSNP:1056602).
 FT /FTId=VAR_011692.
 FT MUTAGEN 255 R->Q: LOSS OF CLEAVAGE.
 FT MUTAGEN 441 S->A: LOSS OF ACTIVITY.
 FT CONFLICT 160 160 M -> V (IN REF. 3).
 FT CONFLICT 242 242 I -> L (IN REF. 1).
 FT CONFLICT 329 329 E -> Q (IN REF. 1).
 FT CONFLICT 489 491 RAD -> KAN (IN REF. 1).
 FT SEQUENCE 492 AA; 53891 MW; CAB44FD174A9076B CRC64;
 Query Match 28.8%; Score 673.5; DB 1; Length 492;
 Best Local Similarity 38.8%; Pred. No. 28-50;
 Matches 149; Conservative 58; Mismatches 128; Indels 49; Gaps 13;
 75 CDGELDCPLGEDEHCVKFPFGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNFTEALA 134
 133 CDGVSHCPGDEENRCVRLY--GP-----NFIQWSSQKSWHPVCODDWNENYG 181
 135 ETACRQMGVSKPTTRAVIGPDQLDVVEITENSQELMRNMSG-----PCL 182
 182 RAACRDMGY--KNNFYSSQ-----GIVDGSGSTSPFKLNTSAGNVDYIKLYHSDACS 232

QY 183 SGLSVLSHLACGKSL---KTPRVVGGEASVDSWPQWQIQYDKQHVCGSILDPHWVL 239
 DB 233 SKAVVSLRCIACGVNLNLSRQSRIVGGSALPGAWPQVSLHVQNHVCGSIITPEWIV 292
 QY 240 TAAHCFRKH-TDVFNVKVRAGSDKLSF---PSLAVAKIIIEFPMY---PKNDIAL 291
 DB 293 TAAHCFRKH-TDVFNVKVRAGSDKLSF---PSLAVAKIIIEFPMY---PKNDIAL 348
 QY 292 MKLQPLTFSGTVRPICLPFDEELTPATLWIGMFTKONGKMSDILLOASVQVIDS 351
 DB 349 MKLQPLTFNDLVKPVCLPNPGLMQLPEQLCSGSGWATEEK-GKTSEVLNAKVLIIET 407
 QY 352 TRCNADDAYQGEVTEKMKACGIEGGVDTGQDSGGPLMYQSDQ-WHVVGIVSGVYGGG 410
 DB 408 QRCNSRYVYDNLITPAMICAGFLQGNVDSCQDSGGPLVTSKNNIWLIGDTSGWGC 467
 QY 411 PSTEGVYTKVSYLNIWYINWYKAE 434
 DB 468 AYREGVYGVNMFVTDWIYRQWRAD 491
 RESULT 5
 TMS3 HUMAN
 ID TMS3_HUMAN STANDARD; PRT; 454 AA.
 AC P57727;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Transmembrane protease, serine 3 (EC 3.4.21.-) (Serine protease
 TAGD-12) (Tumor associated differentially-expressed gene-12 protein).
 GN TMPRSS3 OR TAGD12 OR ECHOS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND T).
 RC TISSUE=Ovarian Carcinoma;
 RX MEDLINE=20521358; PubMed=11069177;
 RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Parmley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel multi-domain cell surface serine
 protease."
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND D), AND VARIANT ILE-53.
 RX MEDLINE=20578749; PubMed=1137999;
 RA Scott H.S., Kudon J., Wattenhofer M., Shibuya K., Berry A., Chraet R.,
 RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
 RA Younis F., Mehdi S.Q., Radhakrishna U., Papasavvas M.P., Gehrig C.,
 RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
 RA Antonarakis S.E.;
 RT "Insertion of beta-satellite repeats identifies a transmembrane
 protease causing both congenital and childhood onset autosomal
 recessive deafness."
 RL Nat. Genet. 27:59-63(2001).
 RN [3]
 RP SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
 RX MEDLINE=2281255; PubMed=12393794;
 RA Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M.,
 RA Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
 RA Buchet K., Raymond A., Hummler E., Marzella P.L., Kudon J.,
 RA Shimizu N., Scott H.S., Antonarakis S.E., Rossier B.C.;
 RT "The transmembrane serine protease (TMPRSS3) mutated in deafness
 DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro."
 RL Hum. Mol. Genet. 11:2829-2836(2002).
 RN [4]
 RP VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.
 RX MEDLINE=21354492; PubMed=11462234;
 RA Masoudi S., Antonarakis S.E., Schwede T., Ghorbel A.M., Gratri M.,
 RA Papasavvas M.P., Drira M., Elgaied-Bouilla A., Wattenhofer M.,
 RA Rossier C., Scott H.S., Ayadi H., Guipponi M.,

RT "Novel missense mutations of TMPRSS3 in two consanguineous Tunisian families with non-syndromic autosomal recessive deafness.";
 Hum. Mutat. 18:101-108(2001).
 [5]
 RN VARIANTS DFNB8/DFNB10 TRP-109; PHE-194 AND ARG-407, AND VARIANTS
 RP ILE-53; SER-111 AND VAL-253.
 RX MEDLINE-21317610; PubMed-11424922;
 RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
 RA Kudoh J., Shibuya K., Antonarakis S.E., Bonne-Tamir B.,
 RA Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
 RA Hance W.E., Wilcox E.R., Friedman T.B., Morell R.J.,
 RA "Novel mutations of TMPRSS3 in four DFNB8/DFNB10 families segregating congenital autosomal recessive deafness.";
 J. Med. Genet. 38:396-400(2001).
 [6]
 RP VARIANTS DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
 RX MEDLINE-21904597; PubMed-11907649;
 RA Wattenhofer M., Di Iorio V., Rabionet R., Dougherty L., Pampanos A.,
 RA Schwede T., Montserrat-Sentis B., Arbones L., Iliades T.,
 RA Pasquadibisceglie A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,
 RA Petersen M.B., Estivill X., Gasparini P., Scott H.S.,
 RA Antonarakis S.E.,
 RA "Mutations in the TMPRSS3 gene are a rare cause of childhood nonsyndromic deafness in Caucasian patients.";
 J. Mol. Med. 80:124-131(2002).
 RL J. Mol. Med. 80:124-131(2002).
 CC -1- FUNCTION: Probable protease. Seems to be capable of activating ENAC.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic reticulum.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Name=A;
 CC IsoId=P57727-1; Sequence=VSP_005391;
 CC Name=B; Synonyms=C;
 CC IsoId=P57727-2; Sequence=VSP_005391;
 CC Name=D;
 CC IsoId=P57727-3; Sequence=VSP_005392;
 CC Name=T; Synonyms=Truncated; TAG-12V;
 CC IsoId=P57727-4; Sequence=VSP_005393;
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues including fetal cochlea. Isoform T is found at increased levels in some carcinomas.
 CC -1- PTM: Undergoes autoproteolytic activation.
 CC -1- DISEASE: Defects in TMPRSS3 are a cause of childhood-onset autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
 CC -1- DISEASE: Defects in TMPRSS3 are a cause of congenital autosomal recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
 CC -1- SIMILARITY: Belongs to peptidase family 11.
 CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF203380; AG37012.1; -
 CC EMBL; AB038157; BAB20077.1; -
 CC EMBL; AB038158; BAB20078.1; -
 CC EMBL; AB038159; BAB20079.1; -
 CC EMBL; AB038160; BAB20080.1; -
 CC HSSP; P00763; 1DPO.
 CC MEROPS; S01.079; -
 CC Genew; HGNC:11877; TMPRSS3.
 CC MIM; 605311; -
 CC MIM; 601072; -
 CC MIM; 605316; -
 CC GO; GO:0016021; C:integral to membrane; NAS.
 CC GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
 CC GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL receptor A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS50287; SRCR_2; 1.
 DR PROSITE; PS02140; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KW Endoplasmic reticulum; Deafness; Alternative splicing;
 KW Disease mutation; Polymorphism.
 FT DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 49 69 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 70 454 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 72 108 LDL-RECEPTOR CLASS A.
 FT DOMAIN 109 205 SRCR.
 FT DOMAIN 217 454 SERINE PROTEASE.
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 401 401 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 216 217 CLEAVAGE (POTENTIAL).
 FT DISULFID 73 85 BY SIMILARITY.
 FT DISULFID 79 98 BY SIMILARITY.
 FT DISULFID 92 107 BY SIMILARITY.
 FT DISULFID 129 194 BY SIMILARITY.
 FT DISULFID 142 204 BY SIMILARITY.
 FT DISULFID 207 324 BY SIMILARITY.
 FT DISULFID 242 258 BY SIMILARITY.
 FT DISULFID 338 407 BY SIMILARITY.
 FT DISULFID 370 386 BY SIMILARITY.
 FT DISULFID 397 425 BY SIMILARITY.
 FT CARBOHYD 221 221 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPPLIC 1 127 Missing (in isoform B).
 FT VARSPPLIC 318 454 /FTId=VSP_005391.
 FT EMBL; AF203380; AG37012.1; -
 FT EMBL; AB038157; BAB20077.1; -
 FT EMBL; AB038158; BAB20078.1; -
 FT EMBL; AB038159; BAB20079.1; -
 FT EMBL; AB038160; BAB20080.1; -
 FT HSSP; P00763; 1DPO.
 FT MEROPS; S01.079; -
 FT Genew; HGNC:11877; TMPRSS3.
 FT MIM; 605311; -
 FT MIM; 601072; -
 FT MIM; 605316; -
 FT GO; GO:0016021; C:integral to membrane; NAS.
 FT GO; GO:0004252; F:serine-type endopeptidase activity; NAS.
 FT GO; GO:0006508; P:proteolysis and peptidolysis; NAS.

```

FT VARIANT 404 404 P -> L (in DFN8/DFN10).
FT VARIANT 407 407 /FTID=VAR_011679.
FT VARIANT 407 407 C -> R (in DFN8/DFN10).
FT VARIANT 426 426 /FTID=VAR_013495.
FT VARIANT 426 426 A -> T.

Query Match 28.2%; Score 660.5; DB 1; Length 454;
Best Local Similarity 35.4%; Pred. NO. 2.4e-49;
Matches 157; Conservative 74; Mismatches 149; Indels 63; Gaps 17;

QY 22 IPMEFRKVGPIITIALSLASIIIVVLKIVLD---KY---YELGQGLHPIPRKQL 74
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 LPLKEFPPIVIGIILIALA-----LGLGHFDCSKYCRSFKC---IELIAR--- 91
QY 75 CDGELDCPLGDEBEHCVKSPPEGPAAVRLSKDRSTLQVLDSATCNWPSACDFNTPALA 134
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 CDGVSDCXGDGEDEYXC-----VRVGQNAVLQVFTAA--SWKTMCSDDWKGHYA 138
QY 135 ETACHOMGYSS-----KPTFRVAVIGPDLDVVEITENSQELMRNNSGCL 182
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 NVACQALGFPPSYVSDNIRVSLQGFREBFVSDIHLLPDDKVTALHHSVTVREG---CA 195
QY 183 SGLSVSLHCLAG--KSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWVILTA 241
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 SGHVVTLQCTACGHRGYSRIVGNGMSLLQWFPQASLQFGYHLCGSVITPLWITA 255
QY 242 AHCFRKHDTVF---NWKVRAGDKL--GSFPLAVAKIIIIIEFNPMY-PK--DNDIALMK 293
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 ARCV---YDLPLPKSWTIQVGLSVLLDNPAPSHLVKIV--YHSKYPRKELGNDIALMK 309
QY 294 LQFPLTFSTVRPICLPFFDELTATPLWIIGWFTKQNGKMSDILLQASVOVIDSTR 353
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 LAGPLTFNEMQVCLPNSSENFPPGKVCWTSGWAGDGDGADGASVPLNHAAPLSNKI 369
QY 354 CNADDAYQGEVTERKMKACIPEGGYDTCCQSGGFLPMYQSPQ-MHWVGIIVSGVCGCGPS 412
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 CNHRDVYGGIISPSMLCAGYLTVGGVDSQSGSGGFLVCQERRLKLKLVGATSGFGICAEVN 429
QY 413 TPGVYTKVSAYLNWYVWKAEL 435
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 KPGVYTRVTSFLDTHIEQWERDL 452

RESULT 6
TWS2 MOUSE STANDARD; PRT; 490 AA.
ID TWS2_MOUSE AC Q9JUK4; Q9JY82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic
DE Transmembrane protein X).
GN TMRPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c;
RX MEDLINE=21104370; PubMed=11169526;
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vinko P.T.;
RT "Expression of transmembrane serine protease TMRPS2 in mouse and
RT human tissues.";
RL J. Pathol. 193:134-140(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;

```

```

RX MEDLINE=20148617; PubMed=10683448;
RA Jacquinet E.J., Rao N.V., Rao G.V., Hoidal J.R.;
RT "Cloning, genomic organization, chromosomal assignment and expression
RT of a novel mosaic serine proteinase: epitheliasin.";
RL FEBS Lett. 468:93-100(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Schuster R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan A., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Kzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
CC cleavage and secreted (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed mainly in prostate and kidney.
CC -1- SIMILARITY: Belongs to peptidase family S1.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF199362; AAF97867.1; -
CC EMBL; AF243500; AAF64186.1; -
CC EMBL; AF113596; AAF21308.1; -
CC EMBL; BC038393; AAF38393.1; -
CC HSSP; P00763; IDPO.
CC MEROPS; S01.247; -.
CC MGD; MGI:1354381; Tmprss2.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00057; ldl_recept_a; 1.
CC Pfam; PF00530; SRCR; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00192; LDLA; 1.
CC SMART; SM00202; SR; 1.
CC SMART; SM00202; Tryp_Spc; 1.
CC PROSITE; PS01209; LDLRA_1; 1.
CC PROSITE; PS00068; LDLRA_2; 1.
CC PROSITE; PS00420; SRCR_1; FALSE_NEG.
CC PROSITE; PS00287; SRCR_2; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT TRANSMEMBRANE PROTEASE, SERINE 2, NON-
FT CATALYTIC CHAIN.

```

FT CHAIN 254 490 TRANSMEMBRANE PROTEASE, SERINE 2,
FT FT CATALYTIC CHAIN.
FT DOMAIN 1 83 CYTOPLASMIC (POTENTIAL).
FT FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 84 104 (POTENTIAL).
FT FT
FT DOMAIN 105 490 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 111 149 LDL-RECEPTOR CLASS A.
FT DOMAIN 150 242 SRCR.
FT DOMAIN 254 490
FT FT SERINE PROTEASE.
FT ACT_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 253 254 CLEAVAGE (POTENTIAL).
FT DISULFID 76 125 BY SIMILARITY.
FT DISULFID 119 138 BY SIMILARITY.
FT DISULFID 132 147 BY SIMILARITY.
FT DISULFID 171 230 BY SIMILARITY.
FT DISULFID 184 240 BY SIMILARITY.
FT DISULFID 243 363 INTERCHAIN (BY SIMILARITY).
FT DISULFID 279 295 BY SIMILARITY.
FT DISULFID 408 424 BY SIMILARITY.
FT DISULFID 435 463 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 474 474 S -> L (IN REF. 3).
FT CONFLICT 122 122 S -> N (IN REF. 3).
FT CONFLICT 178 178 Y -> H (IN REF. 1).
FT CONFLICT 320 320 N -> D (IN REF. 1).
FT CONFLICT 474 474
FT SEQUENCE 490 AA; 53479 MW; 07D2B03EA4D8A19 CRC64;
Query Match 27.7%; Score 648.5; DB 1; Length 490;
Best Local Similarity 35.2%; Pred. No. 2, 9e-48;
Matches 161; Conservative 63; Mismatches 175; Indels 59; Gaps 15;
QY 16 PLKRPIMETKRV-----GIPI-----IALLSLIIIVVILKVIIDKY- 59
DB 50 PQYAPRITQASTVTHTPKSSGAPCTSKSKSLCLALGTVLTAAGAVALLLRFWD 109
QY 60 -----FLCGQLPIPRKQKCDGELDCPEDEHCVKSPPEGFAVAVRLSKDRSTLQ 112
DB 110 SNCSTSEMECGSGTCISSSLWCDGVHCPNGEDENRC-----VRLYGQSFILO 158
QY 113 VLDSATGNWFSACFNFTALAEATACROMYSSKPTFAVEIGPDQ--LDVVEITENSQ 170
DB 159 VYSSORKAWPYCODMSESYGRAACKDMGY--KNFYSSQGIPIQSGATSPFMKLNVSQ 216
QY 171 ELRMNS---SGPCLSGSLVSLHCLACG-KSLK-TPRVVVGGEASVDSWPMQVSTQYDKQ 225
DB 217 NVDLKYLHSDSCSRMVSRHCIEGVRSVRQSRIVUGLNASPGDPWPQVSLHVGQV 276
QY 226 HVCGGSILDPHWLTAACFR-----KHTDVFNKVRAGSKLGSFPFLSAVAKIIIEF 279
DB 277 HVCGGSIIIPETIVTAACVBEPLSGPRYWTAFAGILRQSLMFYGS--RHQVEKVI---S 331
QY 280 NPMY---PKNDIALMKLOPFTSGTVRPICLPFDEBELPATPLWILGWGFTKQNGK 336
DB 332 HENYSKTKNDIALMKLOTPAFNDLVKPVCLPNPGLMDLDQECVSGWGATYK-GK 390
QY 337 MEDILLQASVQVIDTRCNADDAVOGEVTEKMKACAGIPEGGVDTCCQDGGGGLM-YQSDQ 395
DB 391 TSDVLNAAMVPLIEPSKCNKYIYNLITPAMICAGFLQGSVDSQCGSGGGLVTLKNGI 450
QY 396 WHVGVIVSGYCGGSPSPGVTVTKSAVLNLYNWKA 433
DB 451 WWLIGTWSGSCAKALRGVGVNVTFTDIYQOMRA 488
RESULT 7
TMS5_MOUSE
ID TMS5_MOUSE STANDARD; PRT; 455 AA.
AC Q9ER04; Q9ER02; Q9ER03;
DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
GN Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
OS TMPSRS.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBT_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "cDNA cloning of mouse spinesin";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RA Mitsui S., Yamaguchi N.;
RT "Molecular cloning of mouse type 4 spinesin";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- ALTERNATIVE PRODUCTS;
CC Event-Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q9ER04-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9ER04-2; Sequence=VSP_005397, VSP_005398;
CC Name=3;
CC IsoId=Q9ER04-3; Sequence=VSP_005395;
CC Name=4;
CC IsoId=Q9ER04-4; Sequence=VSP_005396;
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB016229; BAB20276.1; -
CC EMBL; AB016230; BAB20277.1; -
CC EMBL; AB016423; BAB20278.1; -
CC EMBL; AB041037; BAB40328.1; -
CC HSP; P00763; IDPO.
CC MEROPS; S01.313; -
CC MCD; MGI.1933407; Tmpress.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYD_SPC; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC PROSITE; PS00420; SRCR_1; FALSE_NEG.
CC PROSITE; PS0287; SRCR_2; 1.
CC Hydrolase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 71 455
FT DOMAIN 112 207 SRCR.
FT DOMAIN 218 455 SERINE PROTEASE.
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).

FT DISULFID 135 196 BY SIMILARITY.
 FT DISULFID 148 206 BY SIMILARITY.
 FT DISULFID 209 328 BY SIMILARITY.
 FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 1 144 Missing (in isoform 2).
 FT VARSPLIC 1 10 /FTID=VSP 005395.
 FT VARSPLIC 1 10 Missing (in isoform 3).
 FT VARSPLIC 1 182 /FTID=VSP 005396.
 FT VARSPLIC 1 182 Missing (in isoform 1).
 FT VARSPLIC 183 192 /FTID=VSP 005397.
 FT VARSPLIC 325 325 /FTID=VSP 005398.
 FT CONFLICT 325 325 D -> G (IN REF. 1; BAB20277).
 SQ SEQUENCE 455 AA; 49632 MW; 5FCF31789C6899AA CRC64;
 Query Match 26.8%; Score 628.5; DB 1; Length 455;
 Best Local Similarity 36.2%; Pred. No. 1.4e-46;
 Matches 135; Conservative 63; Mismatches 140; Indels 35; Gaps 10;
 QY 79 LDCPLGEDEHCVKSPPECPAVAVPLSKDRSTLOVLDSATGNWFSACFDNTEALAEAC 138
 DB 91 LNCPOGVCSEELPLPK--TVSFRINGEDLLQVQRARPDMLLVCHEGSPALGMHIC 148
 QY 139 ROMGYSSKPTFAVPIGPDQDLDVVEITENSCELRMNS-----SGPCLSGSL 186
 DB 149 KSLGHRILQHKAVNL---SDIKL---NRSEPAQLSARPGGLVEAWKPSANCPSGRI 201
 QY 187 VSLHCLACKSLKTRVVGGEASVDSPWQVSIQYDKQVCGGSLDPHWLFAAHC-- 244
 DB 202 VSLKCEGARGPLASRIYGGQAVASGRWFWQASVNLGSRHTCGASVLAHPHWVTAACHMY 261
 QY 245 -PRKHTDVENKVRAGSKLGS---PFLAVAKILIIIEPNMYPKN---DIALMKLQFP 297
 DB 262 SFRL-SRISSEWVHAGLVSHGAVRHOQTWEKIIP---HFLYSQNHQYDVALQLRTP 317
 QY 298 LTFSTVTRPICLPFFDELTLPATPLIIGWFTKONGKMSDILLOASVQVVIDTRCNAD 357
 DB 318 INFSDTDAVCLPAKEQFPWMSQCVSGWGHDTSPSHSTDTLQDTWVPLLSHLNCSS 377
 QY 358 DAYQEVTEKMKACGIPGGVDTCGDSGGPLMYOS-DQWVVGIVSKVCGGSPSTGV 416
 DB 378 CMYSALTRMLCAGYLDGRADACQDSGGLVCPSGDTWHLVGVSWGRGCAEFNRGV 437
 QY 417 YTKVSAYLNWIYN 429
 DB 438 YAKVAEFLDWIHD 450
 ID HEPHS HUMAN STANDARD; PRT; 417 AA.
 AC P05981;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine
 DE 1).
 GN HFN OR TMRSS1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=88209431; PubMed=2835076;
 RA Leytus S.P., Loeb K.R., Hagen P.S., Kurachi K., Davie E.W.;

RT "A novel trypsin-like serine protease (hepsin) with a putative
 RL transmembrane domain expressed by human liver and hepatoma cells.";
 RN Biochemistry 27:1067-1074(1988).
 RC SEQUENCE FROM N.A.
 RP TISSUE=Pancreas, and Spleen;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences".
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
 RL [3]
 RP CHARACTERIZATION.
 RX MEDLINE=91358502; PubMed=1885621;
 RA Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
 RA Chou S.H., Kurachi K.;
 RT "Hepsin, a cell membrane-associated protease. Characterization,
 RT tissue distribution, and gene localization.";
 RL J. Biol. Chem. 266:16948-16953(1991).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=93348237; PubMed=8346233;
 RA Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
 RT "Hepsin, a putative cell-surface serine protease, is required for
 RT mammalian cell growth.";
 CC Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
 CC -!- FUNCTION: Plays an essential role in cell growth and maintenance
 CC of cell morphology.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -!- TISSUE SPECIFICITY: Present in most tissues, with the highest
 CC level in liver.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EXBL; M18930; AAA36013.1; -;
 DR EXBL; X07732; CAA30558.1; -;
 DR EXBL; X07002; CAA30058.1; -;
 DR EXBL; BC025716; AAH25716.1; -;
 DR PIR; S00845; S00845.
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.224; -;
 DR Genew; HGNC:5155; HPN.
 DR MTM; 142440; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008236; F:serine-type peptidase activity; TAS.
 DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR Pfam; PF00089; trypsin; 1.

QY 249 TDVFN-WKVRAGSKLGSFSLAVAKIIIIIBENMYP-----KNDIALMKQLPPLT 299
 Db 228 NRVLNRWRFAGAVARTSPHAFVQLGVQAVIYHGGYLPFRDPTIDENSNDIALVHLSSLP 287
 QY 300 FSGTVRPTCLPFDFDELTPATPLWIGGFTKONGKMSDILLOASVOVIDSTRCNADDA 359
 Db 288 LTEYIOPVCLPAAGALVDGKVCVTGWGNT-QYGGQAMVLOARPIISNEVCNSPDP 346
 QY 360 YQGEVTERKMCAGIPEGGVDFCQDGGGLMYQ-----SDQHWVGVISWYGGCGSPSTP 414
 Db 347 YGNQIKPMFCAGYPEGGIDACQDGGSPFCEDSISGTSWRLCIGVSWGTGCALARKP 406
 QY 415 GVTYKVSAYLWNIYNWK 432
 Db 407 GVTYKVTDFREWIFAK 424

RESULT 11

TMSS5_HUMAN
 ID TMSS5_HUMAN STANDARD; PRT; 457 AA.
 AC Q9H3S3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMPS55.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=11741986;
 RA Yamaguchi N., Okui A., Yamada T., Nakazato H., Mitsui S.;
 RT "Spinesin/TMP55, a novel transmembrane serine protease, cloned from
 human spinal cord.";
 RJ J. Biol. Chem. 277:6806-6812(2002).
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: Brain-specific. Predominantly expressed in
 CC neurons, in their axons, and at the synapses of motoneurons in the
 CC spinal cord.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB028140; BAB20375.1; --
 DR HSSP; P00763; IDPO.
 DR Genew; HGNC:14908; TMPS55.
 DR MIM; 606751; --
 DR MEROPS; S01.313; --
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srcr_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp SPc; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; FALSE_NEG.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein.
 FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT DOMAIN 71 457 (POTENTIAL).
 FT DOMAIN 112 207 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 218 207 SRCR.
 FT ACT_SITE 258 258 SERINE PROTEASE.
 FT ACT_SITE 308 308 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 217 218 CLEAVAGE (POTENTIAL).
 FT DISULFID 135 196 BY SIMILARITY.
 FT DISULFID 148 206 BY SIMILARITY.
 FT DISULFID 209 328 BY SIMILARITY.
 FT DISULFID 243 259 BY SIMILARITY.
 FT DISULFID 374 390 BY SIMILARITY.
 FT DISULFID 401 429 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 457 AA; 49574 MW; 64406AB4985A2851 CRC64;
 Query Match 24.5%; Score 574; DB 1; Length 457;
 Best Local Similarity 32.2%; Pred. No. 7.2e-42;
 Matches 129; Conservative 70; Mismatches 151; Indels 50; Gaps 12;
 QY 60 FLC---GQPLHFTPRKQLCDGE--LDCPLGEDDEHCVKFPPEGPAVAVRLSKDRSTLQVL 114
 Db 71 YLCPAASQPI-----SGTLQDEEITLSCSASAEALLPALPK--TVSFRIINSEDFLLEAQ 124
 QY 115 DSATGNWFSACFDNFTAEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVVEITNSQE--- 171
 Db 125 VRDQPRWLLVCHGEGSPALGLQICWSLGH-----LRLTHKGVNLTDLKLNSSQEFA 176
 QY 172 -----LRMNSSGPCLSGLSVLSLHCLACKSLKTPRVVGGSEASVDSPWQV 218
 Db 177 QLSFRLGGLFEAEAWQPRNN--CTSGVWSLRCSECGARPLASRIVGGOSVAFGRWPQA 233
 QY 219 SIQYDKQHVCGGSIIDPHWVLTAAHCFR--KH'DVFNKVKRAGSDKLGFPSPSLAVAKII 276
 Db 234 SVALGRHTCGGSLVLA PRVWVTAACHWHSFRLARLSRWVHAG--LVSHSAVRPQGAL 290
 QY 277 IEF---NPMYPKDN---DIALMKLPFLTFSGTVRPICLPFFDELTPTATPLMIWGFT 330
 Db 291 VEIIIPLYSAQNDYDVALRLQALNFSDTVGAVCLPAKEHQHPKSGRCWVSWGHT 350
 QY 331 KQNGKMSDILLOASVOVIDSTECNADDAVQGEVTERKMCAGIPEGGVDFCQDGGSP 390
 Db 351 HPHTYSSDMLQDTVVPLSTQLCNSSCVYSGALTFRMLCAGYLDGRADACQDGGSP 410
 QY 391 Y-QSDQHWVGVISWYGGCGSPSTPGVYTKVSAYLWNIYN 429
 Db 411 CPDGDTRWLVGVVSWGRACAEPNHGPYAKVAEFLDWIHD 450
 RESULT 12
 ENTK_BOVIN
 ID ENTK_BOVIN STANDARD; PRT; 1035 AA.
 AC P98072;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTK.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Duoenum;
 RX MEDLINE=94329561; PubMed=8052624;

RA. Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RL protease composed of a distinctive assortment of domains.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RP [2]
RX SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94043122; PubMed=8226855;
RA Lavallie E.R., Rehmetulla A., Racie L.A., Diblasio E.A.,
RA Ferenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase";
RL J. Biol. Chem. 269:23311-23317(1993).
RN [3]
RP SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska H.;
RT "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase";
RL J. Protein Chem. 10:475-480(1991).
CC -!- FUNCTION: Responsible for initiating activation of pancreatic
CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
CC A). It catalyzes the conversion of trypsinogen to trypsin which in
CC turn activates other proenzymes including chymotrypsinogen,
CC procarboxypeptidases, and proelastases.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
CC trypsinogen.
CC -!- SUBUNIT: Heterodimer of a catalytic (light) chain and a
CC multidomain (heavy) chain linked by a disulfide bond.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P98072-1; Sequences=Displayed;
CC Name=Short;
CC IsoId=P98072-2; Sequences=VSP 005386;
CC -!- TISSUE SPECIFICITY: Intestinal brush border.
CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -!- SIMILARITY: Contains 1 SRCR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U09859; AAB40026.1; -;
CC EMBL; L19663; AAA16035.1; -;
CC PIR; A43090; A43090.
CC PDB; 1EXB; 14-OCT-99.
CC MEROPS; S01.156; -;
CC InterPro; IPR000859; CUB.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR002172; LDL_receptor_A.
CC InterPro; IPR000998; MAM domain.
CC InterPro; IPR01254; Peptidase_S1.
CC InterPro; IPR01314; Peptidase_S1A.
CC InterPro; IPR000082; SEA domain.
CC InterPro; IPR001190; Srcr_receptor.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00057; ldl_recept_a; 2.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF01390; SEA; 1.
CC Pfam; PF00530; SRCR; 1.
CC Pfam; PF00089; trypsin; 1.

DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PRINTS; PRO0261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLA; 2.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00200; SEA; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS00060; MAM_2; 1.
DR PROSITE; PS00024; SEA; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Signal-anchor; Glycoprotein; Myristate; Hydrolase; Serine protease;
KW Zymogen; Transmembrane; Repeat; Alternative splicing; 3D-structure;
KW Lipoprotein.
FT CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL)
FT DOMAIN 48 1035 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 54 169 SEA.
FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 240 350 CUB 1.
FT DOMAIN 358 520 MAM.
FT DOMAIN 540 650 CUB 2.
FT DOMAIN 657 695 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 694 787 SRCR.
FT DOMAIN 801 1035 SERINE PROTEASE.
FT ACT_SITE 841 841 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 892 892 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 987 987 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID 2 2 N-myristoyl glycine (Potential).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 236 BY SIMILARITY.
FT DISULFID 659 671 BY SIMILARITY.
FT DISULFID 666 684 BY SIMILARITY.
FT DISULFID 678 693 BY SIMILARITY.
FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
FT DISULFID 826 842 BY SIMILARITY.
FT DISULFID 926 933 BY SIMILARITY.
FT DISULFID 957 972 BY SIMILARITY.
FT DISULFID 983 1011 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 166 Missing (in isoform Short).
FT FTId=VSP_005386.

```

FT CONFLICT 808 808      R -> Y (IN REF. 3)
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

Query Match      24.4%; Score 571; DB 1; Length 1035;
Best Local Similarity 35.0%; Pred. No. 3.9e-41;
Matches 133; Conservative 67; Mismatches 138; Indels 42; Gaps 15;

QY 59 IPRKOLCDGELDCPLGEDEEHCXKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDN 128
DB 672 IPLVNLCDGPHCKDGSDEAHCVRRLF-NGTTDSSGLVQFR--IQSI-----WHVACAEN 722
QY 129 FTEALAEACRQMGV-----SSKPTFRAVEIGPDODLDVVE-----ITENSQBLMRNNSGP 180
DB 723 WTTQISDDVCQLGLGTCNSSVPTF-STGGGPPYVNLNTPNGSLILTPSQ-----772
QY 131 CLSGSLVSLHC--LACGSLKT-----PRVVGGEASVDSWPQVSIQYDKHVCVGGSLD 234
DB 773 CLEDSLILQLQCNYSKCGKLVQEVSPKIVGSDSREGAMPVVALYFDQDQVCGASLVS 832
QY 235 PHWVLTAAHC-FRKHTDVFNKVRAGSKLGSF--PSLAVAKIIIEFNPMY---PKDND 288
DB 833 RDLVLSAAHCYVGRNWSKWKAVLGLHMASNLTSPOIETRLDIQIVINPHYNKRNND 892
QY 289 IALMKLQPLTFSGTVRPICLPFPDEELTPATPLWIGWGTKONGKMSDILLQASVQV 348
DB 893 IAWHLEMKVNYDTYIQPICLPEENQVPPGRCISAGGALTYQ-GSTADVLOEADVPL 951
QY 349 IDSFTCNADDAVQGVTEKMKCAGIPEGGVDTCCGDSGGPLMYQ-SDQWHVVGIVSGYG 407
DB 952 LSNEKQ-QQPEYNITENWVCAGYEAGVDSGCCDSGGLMCCENKRLLAGVTSFYQ 1010
QY 408 CGGSTPGVYTKVSAYLNI 427
DB 1011 CALPNRPGVYARVPRTEWI 1030

RESULT 13
HEPS_RAT      STANDARD;      PRT; 416 AA.
AC Q05511;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serine protease hepsin (EC 3.4.21.-).
GN HPN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93305733; PubMed=8318546;
RA Farley D., Raymond F., Nick H.;
RT "Cloning and sequence analysis of rat hepsin, a cell surface serine
RT proteinase.";
RL Biochim. Biophys. Acta 1173:350-352(1993).
CC -!- FUNCTION: Plays an essential role in cell growth and maintenance
CC of cell morphology.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X70900; CAA50256.1; -.
DR PIR; S33777; S33777.
DR HSSP; P00763; 1DPO.

```

```

DR MEROPS; S01.224; -.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Src receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161
FT CHAIN 162 416
FT CHAIN 416 416
FT DOMAIN 1 16
FT DOMAIN 17 43
FT DOMAIN 44 416
FT DOMAIN 162 416
FT ACT_SITE 202 202
FT ACT_SITE 256 256
FT ACT_SITE 352 352
FT DISULFID 152 276
FT DISULFID 187 203
FT DISULFID 321 337
FT DISULFID 348 380
FT CARBOHYD 111 111
SQ SSSEQUENCE 416 AA; E5A9F8FA950E180 CRC64;

Query Match      24.3%; Score 568.5; DB 1; Length 416;
Best Local Similarity 30.5%; Pred. No. 1.9e-41;
Matches 137; Conservative 64; Mismatches 155; Indels 93; Gaps 11;

QY 19 KPRIPMETRKVGIPILIIALLSLASIIIVVLIKVILDKYFLCGPLHPIPRKQLDGE 78
DB 14 RPXVAALT---VGLTLFTIGGAASWAIITLLR-----44
QY 79 LDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDNTEALAEATC 138
DB 45 -----SDQE-----PLYQVQLSPGDSRLVLVDKTEGTWRLLCSSRNARVAGLC 89
QY 139 ROMGY-----SSKPTFRAVEIG---PPQDLVDVVEITENSQBLMRNSS 178
DB 90 EEMGFLRALAHSELVDVRTAGANGTSGFVCDEGLPLAQRLDLDVISCVD-----138
QY 179 GPLCSGLVSLHLACG-KSLKTPRVVVGGEASVDSWPQVSIQYDKHVCVGGSLDPHW 237
DB 139 --CFRGRFLTATCCDCGRRLKLPVDRIVGGDSGLGRWPQVSLRYDGTLLCGGSLLSGDW 196
QY 238 VLTAACHCFRKHTDVFN-WKVRAGSDKLGPSPSLAVAKIIIEFNPMYP-----KDND 288
DB 197 VLTAACHCFERNVLSRWRFAGAVARTSPHAVQLGVQAVIYHGGVLPFRDPTIDENSND 256
QY 289 IALMKLOPLTFSGTVRPICLPFPDEELTPATPLWIGWGTKONGKMSDILLQASVQV 348
DB 257 IALVHLSSSLPTEYIQVCLPAAGQALVDGKVTYVTGNGT-QFYGQAAVVLQEARVPI 315
QY 349 IDSTRCNADDAVQGVTEKMKCAGIPEGGVDTCCGDSGGPLMYQ-----SDQWHVVGIVS 403
DB 316 ISNEVCNSPDFYGNQIKPMFCAGYPEGGIDACQDSGGHFCVEDRISGTSRWLQGIVS 375
QY 404 WGVCGGSPSTPGVYTKVSAYLNIYNVWK 432
DB 376 WGTGCALARKPGVYTKVIDPFWIFQAIK 404

RESULT 14
HATT_HUMAN
ID HATT_HUMAN
AC O60235;
STANDARD;      PRT; 418 AA.

```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Airway trypsin-like protease precursor (EC 3.4.21.-).
 GN HAT.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=98234382; PubMed=9565616;
 RA Masuda K.-I., Masuda K.-I., Ogawa H., Takagi K., Umemoto N., Yasuoka S.,
 RA "Cloning and characterization of the cDNA for human airway trypsin-
 RT like protease.";
 RL J. Biol. Chem. 273:11895-11901(1998).
 RN [2]
 RN SEQUENCE OF 187-206, AND CHARACTERIZATION.
 RX MEDLINE=97224034; PubMed=9070615;
 RA Yasuoka S., Ohnishi T., Kawano S., Teuchihashi S., Ogawara M.,
 RA Masuda K.-I., Yamakita K., Takahashi M., Sano T.;
 RT "Purification, characterization, and localization of a novel
 RT trypsin-like protease found in the human airway.";
 RL Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).
 CC -!- FUNCTION: May play some biological role in the host defense system
 CC on the mucous membrane independently of or in cooperation with
 CC other substances in airway mucous or bronchial secretions.
 CC -!- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of
 CC arginine residues at the pi position of certain peptides, cleaving
 CC Boc-Phe-Ser-Arg-4-methylcoumaryl-7-amide most efficiently and
 CC having an optimum pH of 8.5 with this substrate.
 CC -!- ENZYME REGULATION: Strongly inhibited by diisopropyl
 CC fluorophosphate, leupeptin, antipain, aprotinin, and soybean
 CC trypsin inhibitor, but hardly inhibited by secretory leukocyte
 CC protease inhibitor at 10 microm.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Activated by
 CC cleavage and secreted.
 CC -!- TISSUE SPECIFICITY: Located in the cells of the submucosal serous
 CC glands of the bronchi and trachea.
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 1 SEA domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: AB002134; BAA28691.1; --
 DR HSSP; P00750; IRTF.
 DR MEROPS; S01.301; --
 DR MIM; 605369; --
 DR GO; GO:0005576; C:extracellular; TAS.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008233; F:peptidase activity; TAS.
 DR GO; GO:0007585; P:respiratory gaseous exchange; TAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001314; Peptidase S1A.
 DR InterPro; IPR000082; SEA_domain.
 DR Pfam; PF01390; SEA; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00200; SEA; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS0024; SEA; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
 KW

Glycoprotein. 1 185 AIRWAY TRYPsin-LIKE PROTEASE, NON-
 CHAIN CATALYTIC CHAIN.
 FT CHAIN 187 418 AIRWAY TRYPsin-LIKE PROTEASE, CATALYTIC
 FT CHAIN CHAIN.
 FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 41 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 42 418 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 44 164 SEA.
 FT DOMAIN 187 417 SERINE PROTEASE.
 FT ACT_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 368 368 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 173 292 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 212 228 BY SIMILARITY.
 FT DISULFID 337 353 BY SIMILARITY.
 FT DISULFID 364 393 BY SIMILARITY.
 FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 418 AA; 46263 MW; F4BCIDB020CFBBD0 CRC64;
 Query Match 24.3%; Score 568; DB 1; Length 418;
 Best Local Similarity 32.4%; Pred. No. 2.1e-41;
 Matches 138; Conservative 79; Mismatches 149; Indels 60; Gaps 14;

QY 33 PIIALLSLASIIIVVLKVIDKYFLCGQPLHPIRK--QLCDGELDCPL----- 83
 DB 16 PYVVCIVVAGVILAVTIALLV---YFLAFDQKSYFYRSSFOLLNVEYNSQLNSPATQE 72
 QY 84 -----GEDEEHCVKSPPEG-----PAVAVRLSKD-----RSTLQVLDGATGNWFSACPD 127
 DB 73 YRTLSGRISLTKTKESNLRNQFIKRAHVAKLRDQSGVRADVMVKFQFTENNGASM- 131
 QY 128 NFEALAEATACROMGYSSKPTTFAVEIGDQDLVDVVEITENSQELMRN--SSGCLSGS 185
 DB 132 ---KSRIEVLRLQNLNS---GNLEINP---STEITSLTDQAAANWLINCEGAGPDLI-- 180
 QY 186 LVSLHCLACGSKLTPRVVVGGEASVDSMPQVSIQYDKQHVCGGSLDPHWLTAACHF 245
 DB 181 -----TLSEORILGTEAEBSWPQVSLRLNNAHCGGSLINNMILTAACHF 229
 QY 246 RKHTDVENKVRAGSKLGSFPSSL--AVAKILIIIEFNPMYKNDIALMKLOFPPTFSCT 303
 DB 230 RSNNSPRDMATSGIST--TFPKLRMRVRNLIHNHYKSATHENDIALVRLNSVTFTKD 287
 QY 304 VRPICLPFFDEELTPATPLWILGWFTKQNGGMSDILLQASVQVITDSTRCHADDAVQGE 363
 DB 288 IHSVCLPAATQNIIPGSTAYVTGWAQAEYAGHTVPE-LRQGGVRIISNDVNCNAPHSYNGA 346
 QY 364 VTEKMMKAGIPEGGVDTCCGDSGGPLMYQSDQ--MHWVGIVSWGVCGGGFPSTPGVYTKVS 421
 DB 347 ILSGMLCAGVPQGVGDACCQDSGGPLVQDSRRLWFIWGVISWGDQCGLPDKPGVYTRVT 406
 QY 422 AYLNWI 427
 DB 407 AYLDWI 412

RESULT 15
 ENTX_MOUSE
 ID ENTX_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Enterocopeptidase (EC 3.4.21.9) (Enterokinase).
 GN PRSS7 OR ENTX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Duoenum;
 RX MEDLINE=98147142; PubMed=9486188;
 RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.:
 RT "Structure of murine enterokinase (enteropeptidase) and expression in
 RT small intestine during development.";
 RL Am. J. Physiol. 274:G342-G349(1998).
 CC -!- FUNCTION: Responsible for initiating activation of pancreatic
 CC proteolytic proenzymes (trypsin, chymotrypsin and carboxypeptidase
 CC A). It catalyzes the conversion of trypsinogen to trypsin which in
 CC turn activates other proenzymes including chymotrypsinogen,
 CC procarboxypeptidases, and proelastases (By similarity).
 CC -!- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys-Ile-7 bond in
 CC trypsinogen.
 CC -!- SUBUNIT: Heterodimer of a catalytic (light) chain and a
 CC multidomain (heavy) chain linked by a disulfide bond (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -!- SIMILARITY: Belongs to peptidase family S1.
 CC -!- SIMILARITY: Contains 2 CUB domains.
 CC -!- SIMILARITY: Contains 2 LDL-receptor class A domains.
 CC -!- SIMILARITY: Contains 1 MAM domain.
 CC -!- SIMILARITY: Contains 1 SEA domain.
 CC -!- SIMILARITY: Contains 1 SRCR domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U73378; AAC37317.1; -.
 CC HSP; Q07954; 1CR8.
 CC MEROPS; S01.156; -.
 CC MGD; MG1.119523; Prss7.
 CC InterPro; IPR000859; CUB.
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR002172; LDL_receptor_A.
 CC InterPro; IPR000998; MAM_domain.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001314; Peptidase_S1A.
 CC InterPro; IPR009020; Protease_inhib.
 CC InterPro; IPR000082; SEA_domain.
 CC InterPro; IPR001190; Srcr_receptor.
 CC Pfam; PF00431; CUB; 2.
 CC Pfam; PF00057; ldl_recept_a; 2.
 CC Pfam; PF00629; MAM; 1.
 CC Pfam; PF01390; SEA; 1.
 CC Pfam; PF00530; SRCR; 1.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC PRINTS; PR00261; LDLRECEPTOR.
 CC PRINTS; PR00020; MAMDOMAIN.
 CC SMART; SM00042; CUB; 2.
 CC SMART; SM00192; LDLA; 2.
 CC SMART; SM00137; MAM; 1.
 CC SMART; SM00200; SEA; 1.
 CC SMART; SM00202; SR; 1.
 CC SMART; SM00020; Tryp_SPC; 1.
 CC PROSITE; PS01180; CUB; 2.
 CC PROSITE; PS01209; LDLA; 1; 2.
 CC PROSITE; PS00681; LDLA; 2; 2.
 CC PROSITE; PS00740; MAM; 1; 1.
 CC PROSITE; PS00060; MAM; 2; 1.
 CC PROSITE; PS00024; SEA; 1.
 CC PROSITE; PS00420; SRCR; 1; FALSE_NEG.
 CC PROSITE; PS00287; SRCR; 2; 1.
 CC PROSITE; PS00240; TRYPSIN_DOM; 1.
 CC PROSITE; PS00134; TRYPSIN_HIS; 1.
 CC PROSITE; PS00135; TRYPSIN_SER; 1.

KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease; Zymogen;
 KW Transmembrane; Repeat.
 FT CHAIN 1 839 NON-CATALYTIC CHAIN (HEAVY CHAIN).
 FT CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
 FT DOMAIN 1 18 CYTOPLASMIC (TYPE-II MEMBRANE PROTEIN)
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 48 1069 EXTRACELLULAR (POTENTIAL).
 FT SEA.
 FT DOMAIN 52 169 SEA.
 FT DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 270 379 CUB 1.
 FT DOMAIN 387 549 MAM.
 FT DOMAIN 569 679 CUB 2.
 FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 723 816 SRCR.
 FT DOMAIN 830 1069 SERINE PROTEASE.
 FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 229 242 BY SIMILARITY.
 FT DISULFID 236 255 BY SIMILARITY.
 FT DISULFID 249 266 BY SIMILARITY.
 FT DISULFID 688 700 BY SIMILARITY.
 FT DISULFID 695 713 BY SIMILARITY.
 FT DISULFID 707 722 BY SIMILARITY.
 FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 859 875 BY SIMILARITY.
 FT DISULFID 959 1027 BY SIMILARITY.
 FT DISULFID 991 1006 BY SIMILARITY.
 FT DISULFID 1017 1045 BY SIMILARITY.
 FT CARBOHYD 147 147 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 197 197 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 380 380 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 579 579 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 675 675 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 727 727 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 770 770 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 791 791 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 897 897 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 936 936 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 999 999 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1069 AA; 118735 MW; E62549E463743CD CRC64;
 Query Match 23.6%; Score 552.5; DB 1; Length 1069;
 Best Local Similarity 32.5%; Pred. No. 1.6e-39;
 Matches 124; Conservative 71; Mismatches 150; Indels 37; Gaps 12;
 QY 69 IPRKQICDGLDCLGEDEHCHVKSPFGEPAVAVRLSKORSTQVLD-SATGNWFACFD 127
 DB 701 IPLGNLCDSYPHCRDGSDEASCVRF-----LNSTRNNGNLVQFNHSHWIAEAE 750
 QY 128 NFEALAEATACROMGYSSKPTFRAVBI---GPDQDLVWEITENSQELMRNRSQGCLSG 184
 DB 751 NWTQISNEVCHLLGLGSANSPISSTGGP-----FVRVQAQNGSLILTSLQCSQD 805
 QY 185 SLVSLHC--LACGSKLT----PRVVGGEASVDSPWQVSIQYDKQH-----VCGSGILD 234
 DB 806 SLILLQCNHKSCEKXKVTQKVSFKVGGSDAQAGAPWVVALYHRDRSTDRLLCGASLVS 865
 QY 235 PHVVLTAHC-FRKHTDVFNNKVRAGSKLGSPFSLAVAKII--TEFPMYP---KDND 288
 DB 866 SDMLVSAHCHVYRNLDPTRTWTAVLGHMQSNSTPQVVRVVDQIVINHYDRRKVND 925
 QY 289 IALMKIQPLTFSGTVRPICLPFFDEBELTPATPLWIGWFTKQNGGKMSDILLQASVQV 348
 DB 926 IAWHLEFKVNYTDYIQPICLPEENQIFIGRTCSAGWGYDKINAGSTVDVLKEADVPL 995
 QY 349 IDSTRCNADDAYQGEVTEKMKACIGEGGVDTCQGSQGGPLMYQ-SDQWHVGVISWGYG 407

Db 986 ISNERCQ-QQLPEYNITESMICAGYEEGGIDSCQDGGPLMCOENNEWFLVGVTSFGVQ 1044
Qy 408 CCGPSTPGVYTKVSAYLWYN 429
Db 1045 CALPNHPGVYVRSQFIEWHS 1066

Search completed: June 7, 2004, 08:39:20
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:38:47 ; Search time 45 Seconds
(without alignments)
3050.009 Million cell updates/sec

Title: US-10-030-688-2

Perfect score: 2342
Sequence: 1 MDPSPDPLNSLDVYKPLRKP.....VYTKVSAYLWVYNNVWKAEL 435

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_virus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2167	92.5	405	Q96E86	Q96E86 homo sapien
2	704.5	30.1	471	Q8CFE0	Q8CFE0 mus musculus
3	695.5	29.7	581	Q8BYE2	Q8BYE2 homo sapien
4	688	29.4	453	Q81ZA6	Q81ZA6 mus musculus
5	686.5	29.3	537	Q8BYE1	Q8BYE1 homo sapien
6	686.5	29.3	558	Q8EYM4	Q8EYM4 homo sapien
7	686.5	29.3	767	Q9DGR2	Q9DGR2 xenopus lae
8	670.5	28.6	492	Q86T73	Q86T73 homo sapien
9	648	27.7	490	Q7TN04	Q7TN04 mus musculus
10	646	27.6	490	Q920K3	Q920K3 rattus norv
11	636.5	27.2	445	Q8CJ17	Q8CJ17 rattus norv
12	631.5	27.0	455	Q8CDP0	Q8CDP0 mus musculus
13	612	26.1	371	Q8CJ16	Q8CJ16 rattus norv
14	545	23.3	777	Q8CAN9	Q8CAN9 mus musculus
15	536	22.9	417	Q8VJ74	Q8VJ74 rattus norv
16	527	22.5	279	Q9QZ74	Q9QZ74 rattus norv

17	523	22.3	417	11	Q8VDV1	Q8vdv1 mus musculus
18	523	22.3	417	11	Q8VHK8	Q8vhk8 mus musculus
19	521.5	22.3	279	11	Q7TNX3	Q7tnx3 mus musculus
20	510	21.8	328	11	Q8OZ40	Q8oz40 rattus norv
21	504.5	21.5	329	6	Q9GL10	Q9gl10 ovis aries
22	504.5	21.5	1524	13	Q91674	Q91674 xenopus lae
23	503.5	21.5	336	11	Q8OYD8	Q8oyd8 mus musculus
24	502	21.4	638	11	Q8ROP5	Q8rop5 mus musculus
25	501.5	21.4	439	11	Q8BHM9	Q8bhm9 mus musculus
26	501.5	21.4	556	13	Q8O3D5	Q8o3d5 brachydanio
27	500.5	21.4	1111	11	Q8OYN4	Q8oyn4 rattus norv
28	499.5	21.3	327	4	Q8N171	Q8n171 homo sapien
29	499.5	21.3	572	11	Q8BIK6	Q8bik6 mus musculus
30	498.5	21.3	377	6	P79343	P79343 bos taurus
31	498.5	21.3	417	11	Q8BZ10	Q8bz10 mus musculus
32	496	21.2	643	6	Q97506	Q97506 sus scrofa
33	493.5	21.1	421	11	Q6O491	Q6o491 cavia porce
34	493	21.1	416	4	Q86T26	Q86t26 homo sapien
35	488	20.8	423	11	Q8EM10	Q8em10 mus musculus
36	488	20.8	812	11	Q9ROW3	Q9row3 rattus norv
37	486.5	20.8	416	11	Q8BZ30	Q8bz30 mus musculus
38	486.5	20.8	416	11	Q8BZ13	Q8bz13 mus musculus
39	486	20.7	389	13	Q9PVX7	Q9pvx7 xenopus lae
40	485.5	20.7	326	13	Q7ZZ80	Q7zz80 brachydanio
41	482.5	20.6	284	4	Q8NPF6	Q8nfp6 homo sapien
42	481	20.5	321	4	Q96RZ8	Q96rz8 homo sapien
43	479.5	20.5	845	13	Q9DGR1	Q9dgr1 xenopus lae
44	478.5	20.4	415	6	Q29015	Q29015 sus sp. pre
45	477	20.4	1379	5	Q9V4N6	Q9v4n6 drosophila

ALIGNMENTS

RESULT 1
Q96E86
ID Q96E86 PRELIMINARY; PRT; 405 AA.
AC Q96E86;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to transmembrane protease, serine 4 (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; BC012752; AAH12752.1; -
DR HSSP; P00761; IAN1.
DR MEROPS; S01.034; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SMC0192; LDLa; 1.
DR SMART; SMC0202; SR; 1.
DR SMART; SMC0020; Tryp_Spc; 1.
DR PROSITE; PS50287; SRCF_2; 1.

DR PROSITE; PS50240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

FW Hydrolase; Protease; Serine protease; Transmembrane.

KT NON_TER 1

SQ SEQUENCE 405 AA; 44474 MW; 951ACD529D48E04 CRC64;

Query Match 92.5%; Score 2167; DB 4; Length 405;

Best Local Similarity 100.0%; Pred. No. 8.3e-202;

Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 IIALSLASIIIVVILKVIDKYFLCGPLHPIPRKQJCDGELDCPLGEDSEHCVK 93

DB 4 IIALSLASIIIVVILKVIDKYFLCGPLHPIPRKQJCDGELDCPLGEDSEHCVK 63

QY 94 PEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDFNTEALAEACRQGYSSKPTFRAVE 153

DB 64 PEGPAVAVRLSKDRSTLQVLDSATGNWFSACFDFNTEALAEACRQGYSSKPTFRAVE 123

QY 154 IGPDDLDVVEITENSQELRMNNSGPGCLSGSLVSLHCLACGKSLKTPRVVGGSEASVDS 213

DB 124 IGPDDLDVVEITENSQELRMNNSGPGCLSGSLVSLHCLACGKSLKTPRVVGGSEASVDS 183

QY 214 WFWQVSIQYDKOHVCGGSLDPHVLTAHCFRKHDTVFNKVRAGSKLGSFSLAVAK 273

DB 184 WFWQVSIQYDKOHVCGGSLDPHVLTAHCFRKHDTVFNKVRAGSKLGSFSLAVAK 243

QY 274 IIIIEFNPMYKNDIALMKLOFPLTFSGTVRPICLPFDEBELPATPLWIIGGFTKQN 333

DB 244 IIIIEFNPMYKNDIALMKLOFPLTFSGTVRPICLPFDEBELPATPLWIIGGFTKQN 303

QY 334 GSKMSDILLOASVQVIDTRCNADDAVGEVTEKMWKAGIPGGVDTCQDGGGPLYMQS 393

DB 304 GSKMSDILLOASVQVIDTRCNADDAVGEVTEKMWKAGIPGGVDTCQDGGGPLYMQS 363

QY 394 DQWVVGVISWGYGCGGPGSTPGVYTKVSAIYNWVMAEL 435

DB 364 DQWVVGVISWGYGCGGPGSTPGVYTKVSAIYNWVMAEL 405

RESULT 2

Q8CFEO

ID Q8CFEO PRELIMINARY; PRT; 471 AA.

AC Q8CFEO;

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Similar to mosaic serine protease (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Breast tumor;

RA Strausberg R.;

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC042878; AAH42878.1; -

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:004263; F:chymotrypsin activity; IEA.

DR GO; GO:008233; F:peptidase activity; IEA.

DR GO; GO:0005044; F:scavenger receptor activity; IEA.

DR GO; GO:0004295; F:trypsin activity; IEA.

DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.

DR InterPro; IPR009003; Cys Ser trypsin.

DR InterPro; IPR002172; Ldl receptor A.

DR InterPro; IPR001254; Peptidase S1_

DR InterPro; IPR001314; Peptidase_S1A

DR Pfam; PF00057; ldl_recept_a; 1.

DR Pfam; PF00530; SRCR; 1.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR

DR	InterPro; IPR009003; Cys_Ser_trypsin.
DR	Interpro; IPR002172; LDL_receptor A.
DR	Interpro; IPR001254; Peptidase S1.
DR	InterPro; IPR001314; Peptidase S1A.
DR	InterPro; IPR001190; Srcr receptor.
DR	Pfam; PF00057; ldl_recept_a; 1.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	SMART; SM00192; LDLa; 1.
DR	SMART; SM00202; SR; 1.
DR	SMART; SM00020; Tryp Spc; 1.
DR	PROSITE; PS0287; SECR_2; 1.
DR	PROSITE; PSS0240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
KW	Hydrolase; Protease; Serine protease.
SQ	SEQUENCE 581 AA; 62689 MW; 4DABE247D5BA4A4 CRC64;

Query Match	29.7%; Score 695.5; DB 4; Length 581;
Best Local Similarity	35.4%; Pred. No. 1.5e-58;
Matches 152; Conservative 69; Mismatches 161; Indels 47; Gaps 11;	

Qy	22 IPMTFFR--KVGIPITIIALLSLASIIIVVILDKYVFELGG-----PLH 67
Dd	149 LPKTWEQGQQLPGLICVLLIALLVLSLIL-----FQWQHGTHRHKEQRSECPKH 202
Qy	68 FIPRKQLCDGELDCPLGEDEBHCVKSPPEGPAAVRLSKDRSTLQVLDSATGNWFPSACFD 127
Dd	203 AV----RCDGVVDCXKSDELGC-----VRFDWDKSLLIKYSGSSHQWLPICSS 247
Qy	128 NFTAALAEACRMGYGYSKPTFRAVEIGDPDLDVEITEENSQELMRNMSGCLSGSLV 187
Dd	248 NWNDSYSEKTRCQLGFES--AHRTEVAHRDFANFSILRYNSTIQESLHRSHCPSQRYI 305
Qy	188 SLHCLACKSLKTPRVVGGEEASVDSPWQVSIQYDKOHCGGSTLDPHWLTAAHCF-- 245
Dd	306 SLQCSHCLRAMTGEIVGGALASDKWPQVSLHFGTTHICGTLIDAQWLTAAHCEV 365
Qy	246 RKHTDVFNWKVRAGSDKLGSPPSLA-VAKIIIIIEFPNMY---PKNDIALMKLOFPPLTF 301
Dd	366 TREKVLEGWKKYYAGTSNLHLQLPAAASIAEIII---NSNYTDEEDDYDIALMRLSKPLT 422
Qy	302 GTVPICLPFFEDELTAPATPWITGMGTQKNGKGMSDILLQAASVOVIDSTRCNADDAAYO 361
Dd	423 AHIIHPACLPMHGQTFLSNETCMWTGFGKTRTDDTKTSFPLREVQVNLIDPFKKNDLYVD 482
Qy	362 GEVTEKMCACGIPPEGGVDTCCGDSSGPLM-YQSODHWVVGIVSWGYGCGGSPGPGVTYKV 420
Dd	483 SYLIFPRMWCAGDLHGRRDSCQDGGPLVCQNNEWLAGVTSNGTGCGQRKKGVTYKV 542
Qy	421 SAYLNWIYN 429
Dd	543 TEVLPMWYS 551

```

RESULT 4
Q812A6 PRELIMINARY; PRT; 453 AA.
ID Q812A6
AC Q812A6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transmembrane proteinase tmpr33.
DE Mus musculus (Mouse).
OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCEI_taxid=10090;
[1]
RN SEQUENCE FROM N.A.
RP Rao N.V., Rao G.N., Hoidal J.R.;
RA "Genomic Organization of Murine Transmembrane Proteinases.";
RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF479687; AAC33581.1; -.
DR

```

[illegible]

RESULT 5	
Q9BYE1	
ID	Q9BYE1
AC	PRELIMINARY; PRT; 537 AA.
DT	Q9BYE1;
DT	01-JUN-2001 (trEMBLrel. 17, Created)
DT	01-JUN-2003 (trEMBLrel. 17, Last sequence update)
DT	01-OCT-2003 (trEMBLrel. 25, Last annotation update)
DE	Mosaic serine protease.
GN	MSPS.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=21167393; PubMed=11267681;
 RA Kim D.R., Sharmin S., Inoue M., Kido H.;
 RT "Cloning and expression of novel mosaic serine proteases with and
 RL without a transmembrane domain from human lung.";
 RL Biochim. Biophys. Acta 1518:204-209 (2001).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB048797; BAB39742.1; -;
 DR HSSP; P00763; LDPO.
 DR MEROPS; S01.087; -;
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srrc_receptor.
 DR Pfam; PF00057; lcl_receptor; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0287; SRR2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 537 AA; 58102 MW; A39FF4E8816DAECF CRC64;

Query Match 29.3%; Score 686.5; DB 4; Length 537;

Best Local Similarity 38.6%; Pred. No. 18-57; Matches 142; Conservative 56; Mismatches 147; Indels 23; Gaps 7;

QY 75 CDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSDATGNWFSACDFNTEALA 134
 DB 181 CDGVVCKLKSDELGC-----VRFWDKSLKLYSGSSHQWLPICSSNNWDSYS 229
 QY 135 ETACQMGVSSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGCPCLSGSLVSLHCLAC 194
 DB 230 EKTCTQLGPES--AHRTEVAHRDFANSFSLRYNSTIQESLHRSCHSPSRYISLQCSHC 287
 QY 195 GSKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSLDHPHVLTAACHF--RKHTDVF 252
 DB 288 GURANTGRIVGGALASDSKWPQVSLHFGTHICGGTLIDAQWVLTAAHCFVTRKYLE 347
 QY 253 NKKVRAGSKLGSFPFSLA-VAKIIIEFNPMY---PKNDIALMKQLPPLTFSGTVRPI 308
 DB 348 GKKVAGTSLNHLQPEAASIAEII---NSNYTDEDDYDIALMRLSKPLTSLAHIPAC 404
 QY 309 LPFFDEELTPATPLTWIGGFTKQNGKMSDILLQASVQVIDSTECNADDAVQGVTEKM 368
 DB 405 LPMHGQTFSLNCTWTGFKTRTDDKTSPLREVQVNLIDFKCNLYLVDSYLTPRM 464
 QY 369 MCAGIPEGGVDTCCGDSGGPLM-YQSDQHVVGIVSWGCGGSPSTPGVYTKVSAYLNI 427
 DB 465 MCAGDLHGGRDSCQDSGGPLVCEQNNRWLAGVTSWGTGCGQRNKPVGVTYKTVLFWI 524
 QY 428 YNVWAEEL 435
 DB 525 YSRMESEV 532

RESULT 6

Q86YM4 PRELIMINARY; PRT; 558 AA.

AC Q86YM4;
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Transmembrane protease serine 6.
 GN TMPSR56.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Park T.J., Park W.J.
 RT "Homo sapiens transmembrane protease, serine 6 (TMPSR56) mRNA.";
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY190317; AAC38062.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Srrc_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0287; SRR2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Protease.
 SQ SEQUENCE 558 AA; 60432 MW; 4AC817FCD70D7017 CRC64;

Query Match 29.3%; Score 686.5; DB 4; Length 558;

Best Local Similarity 35.1%; Pred. No. 11e-57; Matches 149; Conservative 74; Mismatches 163; Indels 39; Gaps 11;

QY 22 IPMETPR--KVGPIIIALLSLASIIIVVVLKIVLDKYYFLCGQP-LHFIKQLC--- 75
 DB 149 LPRFTWREGQKQLFLGCVLLIALVLSLIL-----PQFWQGYTGIRYKEQRESCEPH 202
 QY 76 ----DGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDSDATGNWFSACDFNTE 131
 DB 203 AVRDDGVVCKLKSDELGC-----VRFWDKSLKLYSGSSHQWLPICSSNNWMD 251
 QY 132 ALAETACQMGVSSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGCPCLSGSLVSLHC 191
 DB 252 SYSEKTCQLGPES--AHRTEVAHRDFANSFSLRYNSTIQESLHRSCHSPSRYISLQC 309
 QY 192 LAGKSLKTPRVVGGEASVDSWPQVSIQYDKQHVCGGSLDHPHVLTAACHF--RKHT 249
 DB 310 SHCGLRAMTGRIVGGALASDSKWPQVSLHFGTHICGGTLIDAQWVLTAAHCFVTRK 369
 QY 250 DVFNKVRAGSKLGSFPFSLA-VAKIIIEFNPMY---PKNDIALMKQLPPLTFSGTVR 305
 DB 370 VLEGKVVAGTSLNHLQPEAASIAEII---NSNYTDEDDYDIALMRLSKPLTSLAH 426
 QY 306 PICLPFDEELTPATPLTWIGGFTKQNGKMSDILLQASVQVIDSTECNADDAVQGVTE 365
 DB 427 PACLPMHGQTFSLNCTWTGFKTRTDDKTSPLREVQVNLIDFKCNLYLVDSYLT 486
 QY 366 EKMWAGIPEGGVDTCCGDSGGPLM-YQSDQHVVGIVSWGCGGSPSTPGVYTKVSAYL 424
 DB 487 PRMWCAGDLHGGRDSCQDSGGPLVCEQNNRWLAGVTSWGTGCGQRNKPVGVTYKTVL 546
 QY 425 NWIYN 429
 DB 547 PWIYS 551

RESULT 7

Q9DGR2 PRELIMINARY; PRT; 767 AA.
 ID Q9DGR2

Q9DGR2;
01-MAR-2001 (TrEMBLrel. 16, Created)
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Embryonic serine protease-2.
GN XESP-2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363741; PubMed=10903452;
RA Yamada K., Takabatake T., Takeshima K.;
RT Isolation and characterization of three novel serine protease genes
from Xenopus laevis.
RL Gene 252:209-216(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AB038497; BAB08217.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.049; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLA; 8.
DR SMART; SM00020; tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 8.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE596E CRC64;

Query Match 29.3%; Score 686.5; DB 13; Length 767;
Best Local Similarity 38.3%; Pred. No. 1.7e-57;
Matches 145; Conservative 58; Mismatches 153; Indels 23; Gaps 7;

QY 62 CQOPLHFIPRKQLCGELDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLQVLDSATGNW 121
DB 395 CGSSVSCVLSQWCGVSDCPYGEDEMSCVSLPAD-----PQLQVYSTSVSAW 443

QY 122 FSACPDNFTALAEATACRMGYSSKPTFRAVEI-----GPDQDLVDVEITENSQELMRNS 177
DB 444 LPVCSDYNDPGRACQDFGNGSSYNYRDTLMSYPAPNGYFKLYSGYWSKFTYVQY 503

QY 178 SGPCLSGSLVSLHCLACGKSLKT--PRVVGGEASVDSWPQVSIQYDKQHVCGSILDP 235
DB 504 SSYCVSGNVSLHCLISCGVSNLSRVISVGGTFANLGNWQVNLQYITGVLCGGSIIIP 563

QY 236 HWLTAACH-FRKHTDVENVKVRAGSKLGSPSLA---VAKIIIEFNPMYPKNDIAL 291
DB 564 KWIVTAACHVGSYSASGWRVAGTLTKPSYINASAYFVERIIVHPGYKYTYNDIAL 623

QY 292 MKLQPLTFSGTVRPICLPFDEBELTPATPLWIGFTKQNGKMSDILLOASVQVIDS 351
DB 624 MKLRDEITFGYTTQVCLPNSGMFEAGTTTWISGWSGTYB-GGSVSTYLYAAIPLIDS 682

QY 352 TRCNADDAAYQEVTEKMKAGIPGGVDTCCGDSGGLM-YQSQWHVGVISWGYCGGG 410
DB 693 NVNCSYVNGGQITSSMTCAGYLSGGVDTCCGDSGGLNKNRGTWMLVGTSTWGDGOAR 742

QY 411 PSTPGVYTKVSAYLNWIYN 429

Db 743 ANKPGVYGVNTTFLEWIS 761
RESULT 8
Q96T73 PRELIMINARY; PRT; 492 AA.
ID Q96T73
AC Q96T73; PubMed=10903452;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Epitheliasin.
DB TMPSR2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21232025; PubMed=11328890;
RA Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,
RA Hoidal J.R.;
RT "Cloning and characterization of the cDNA and gene for human
epitheliasin."
RL Eur. J. Biochem. 268:2687-2699(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AF329454; AAK53559.1; -.
DR HSSP; P00761; IAN1.
DR GO; GO:0015020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR00134; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00202; SR; 1.
DR PROSITE; PS01209; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 492 AA; 53863 MW; 3ABA755BF276DADF CRC64;

Query Match 28.6%; Score 670.5; DB 4; Length 492;
Best Local Similarity 38.8%; Pred. No. 3.3e-56;
Matches 148; Conservative 59; Mismatches 128; Indels 49; Gaps 13;

QY 75 CDGELDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLQVLDSATGNWFSACPDNFTAL 134
DB 133 CDGVSHCPGGEDEHRCVRLY--GP-----NFIQWVSSQKSWHPVCQDDWENY 181

QY 135 ETACRQWYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSG-----PCL 182
DB 182 RAACKQNGY--KNFYSSQ-----GIVDSSGSTSFMKLNTSAGNVDIYKLYHSDACS 232

QY 183 SGLSVSLHCLACGKSL---KTPRVVGGEASVDSWPQVSIQYDKQHVCGSILDPHWVL 239
DB 233 SKAVSLRCLACGVLNLSRQSIIVGESALPGAWPQVSLHVQNVHVCVCGSIIITEWIV 292

QY 240 TAAHCFRKH-TDVFNKKVRAGSKLGSF---PSLAVAKIIIEFNPMY---PKNDIAL 291
DB 293 TAAHCFVEKPLNPNWHTAFAGILR-QSFMFYGAGYQVEKVI---SHPNYDSKTKNNDIAL 348

QY 292 MKQLPFLTSGTVTRPCLPFFDEELTPATPLWIGWFTKONGKMSDILLOASVQVIDS 351
 DB 349 MKQLPFLTNDLVKVPCLPNCMLQPEQLCWSGATEEK-GKTSVLNAKVLIIET 407
 QY 352 TFCNADDAVQGEVTEKWCAGTPEGGVTCCGDSGGLMYQSDQ-WHVVGIVSVNGYCGG 410
 DB 408 QRCNRYVYDNLITPAMICAGFLQGNVDSGDSGGLVTSKNNIWLIGTWSGCGAK 467
 QY 411 PSTPGVYTKVSAYLNWYVWKA 434
 DB 468 AYRPGVYGVWVFTDWIYRQEAD 491

RESULT 9
 Q7TN04 PRELIMINARY; PRT; 490 AA.
 AC Q7TN04;
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Tmprss2 protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129; TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buerow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano P.J., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalek A., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Maria M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129; TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC054348; AAHS4348.1; -;
 SQ SEQUENCE 490 AA; 53525 MW; 54650B028417665A CRC64;

Query Match 27.7%; Score 648; DB 11; Length 490;
 Best Local Similarity 37.1%; Pred. No. 5e-54;
 Matches 155; Conservative 61; Mismatches 162; Indels 40; Gaps 14;

QY 39 LSLASIIIVLVKIVLDKY-----FLGQPLHPIPRKQLCDGELDPLGDEEHC 90
 DB 88 LALGVLTAAGAAVALLRWFDSNCSSTSEMECGSGTCLSSSLWCDGVAHCPNGDENRC 147
 QY 91 VKSFPEGPAVAVRLSKDRSTLQVLDSATCNWFSACFDNFEALAEATACRQMGYSKPTFR 150
 DB 148 -----VRLYQSILQVYSSQKAWPVQODDWSSEYSGRAACKDMGY--KKNFY 194
 QY 151 AVEIGPDQD--LDVVEITNSOELMRNS---SGCLSGSLVSLHCLAGC-KSLK-TPRV 203
 DB 195 SSGQIPDQSGATSFMKLVNSSGVDLYKKLYHSDSCSRMVMVSLRCIECGVRSVKQSGRI 254

QY 204 VGGEASVDSWPNQVSIQYDKQHVCGSILDPHWLTRAHCFRKH-TDYFNWVKVAG--S 260
 DB 255 VGLNASPGDMPQVSLHVQGVHVCVGSIIITPEWIVTAAHCVCEELPSSPRYWTAFAGILR 314
 QY 261 DKLGSPSL-AVAKIIIEFNPMY---PKNDIALMKLOFPLTFSGTVPICLPFFDEEL 316
 DB 315 QSLMFVGSRHOVEKVI---SHPNYDSKTKNDIALMKLQTLPLAFNDLVKPVCLPMPGWL 371
 QY 317 TPATPLWIIWGFTKONGKMSDILLOASVQVIDSTRCNADDAVQGEVTEKMKCAGIPEG 376
 DB 372 DLQECWISGWGATYEK-GKTSVLNAAMVPLIEPSKCNKSIYNNLITPAMICAGFLQG 430
 QY 377 GVTCCGDSGGLM-YQSDQWVHVGVSVNGYCGGSPSTEGVYTKVSAYLNWYVWKA 433
 DB 431 SVDSGDSGGLVTLKNGIWLIGTWSGCGAKALRFGVYGVNTVFTDWIYQOMRA 488

RESULT 10
 Q920K3 PRELIMINARY; PRT; 490 AA.
 AC Q920K3;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Tmprss2.
 GN Tmprss2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tsuzuki S.;
 RT "Tmprss2, Rat."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB073550; BAB70883.1; -;
 DR HSP; P00761; IAN1.
 DR GO; GO:0016020; C.membrane; IEA.
 DR GO; GO:0004263; F.cytotrypsin activity; IEA.
 DR GO; GO:0008233; F.peptidase activity; IEA.
 DR GO; GO:0005044; F.scavenger receptor activity; IEA.
 DR GO; GO:0004295; F.trypsin activity; IEA.
 DR GO; GO:0006508; P.proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001314; Peptidase_S1A.
 DR InterPro; IPR001190; Sckr_receptor.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00068; LDLRA_2; 1.
 DR PROSITE; PS00287; SCKR_2; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 490 AA; 53515 MW; 2BC691551CAC409A CRC64;

Query Match 27.6%; Score 646; DB 11; Length 490;
 Best Local Similarity 38.8%; Pred. No. 7.9e-54;
 Matches 151; Conservative 54; Mismatches 148; Indels 36; Gaps 13;

QY 62 CGOPLHPIPRKQLCDGELDPLGDEEHCVKSPFEPGPAVAVRLSKDRSTLQVLDSATGNW 121
 DB 119 CGSSGTCLSSSLWCDGVAQCPNGKDNRC-----VRLYGSFTLQVYSSQKAW 167
 QY 122 FSACFDNFTALAEATACRQMGYSKPTFAVEIGPDQ-----DLVVEITNSOEL-RMRN 176
 DB 168 YPVQDDWNSYGRAACKDMGY--KNSFYSSQGIPOQSGATSPKLVNVSAGNVLDLYKLY 225

QY 177 SSGPCLSGSLVSLHCLACG-KSL-KTPRVVGGEEASVDSWPKVSIQYDKQHVCGSILD 234
 Db 226 HSDSCSSRWVSLRCEICGVRSVRQSRIVGSGTASPDGDPWQVSLHVQGIHVCGSILT 285
 QY 235 PHWLTAACHCRKFK-TDVFNWKKVRAGSDK-----LGSPFSLAVAKIIIEFNPMYK 285
 Db 286 PEWIVTARHCVEEPLSSPRYWTAFAGILKQSLMFYGS--RHQVEKVI---SHPNVDSKTK 340
 QY 286 DNDIALMKLOPLTSGTVRPTCLIPFDEELTPATPLMIIGWFTKQNGKMSDILLQAS 345
 Db 341 NNDIALMKLOPLTAYDNDVVVKVCLPNPGWMLDLAQECWISGWGATYEK-GKTSVDVLAAM 399
 QY 346 VOVIDSTRCNADDAQGEVTEKMCAGIPGEGVDTCQDSDGGLM-YOSDQWVGVISW 404
 Db 400 VPLIEPSKNSKYIYNLITAMICAGFLQSDVSDSCQSDSGGLVTLKNEIWLWIGTISW 459
 QY 405 GYGCGGSPTPGYTKVSAYLWNIYNWKA 433
 Db 460 GSGCAKAYRPGVYGNVFTDWTIYQOMRA 488

RESULT 11

Q8CJ17 Q8CDRO PRELIMINARY; PRT; 445 AA.
 AC Q8CJ17;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Adrenal mitochondrial protease long variant.
 GN AMP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEBH;
 RA Omer S., Bicknell A.B., Lowry P.J.;
 RT "Identification of a rat adrenal mitochondrial protease."
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF537098; AA06757.1; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001354; Peptidase_S1.
 DR InterPro; IPR001190; Srct_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Protease.
 SQ SEQUENCE 445 AA; 48440 MW; BE3F56D8372ED988 CRC64;

Query Match 27.2%; Score 636.5; DB 11; Length 445;
 Best Local Similarity 36.2%; Pred. No. 5.8e-53;
 Matches 135; Conservative 62; Mismatches 141; Indels 35; Gaps 10;

QY 79 LDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTEALAEATAC 138
 Db 81 LSCPGVSSEKLLPSLPK-AVSRFRNGEDLLLEVQVREARPDLLVCHEGNPAIGWHIC 138
 QY 139 RQMGYSKPTFRAVEIGDPQDLVDVEITENSQELMRNS-----SGPCLSGSL 186
 Db 139 QSLGFRLTQKAVNL-----SDIKL-----NRSQEFALQSARPGSLVEBAWQSTNCPGRI 191

QY 187 VSLHCLACGSLKLTPTRVVGGEEASVDSWPKVSIQYDKQHVCGSILDPHWLTAACHC-- 244
 Db 192 VSLKCEGCGARPASRIVGGQAVASGRWPMQASVILGSRHTCGSVLAPYVVVTAACHMY 251
 QY 245 -FRKHTDVFENWKKVRAG---SDKLGSFPFLSAVAKIIIEFNPMYKPN---DIALMKLQFP 297
 Db 252 SFRL-SRLSSMRVHAGLVSHSAVRQHGTWVEKIIP---HPLYSANQHDYDVALQLRTP 307
 QY 298 LTSSTGVTRPCLIPFDEELTPATPLMIIGWFTKQNGKMSDILLQASVOVIDSTRCNAD 357
 Db 308 INFSDTVSAYCLPKEQHFPGSQCCVSGWGHGTHDPSHTSHSSDLQDTMVPELLSTDLNCS 367
 QY 358 DAYQGEVTEKMCAGIPGEGVDTCQDSDGGLMVQS-DQWVVGVISWGVCGGSPSTPGV 416
 Db 368 CMTSGALTRMLCAGYLDGRADACQSDSGGLVCPSGDTWHLVGVVSWRGCEAPNRPGV 427
 QY 417 YTKVSAYLWNIYN 429
 Db 428 YAKVAEFLDWIHD 440

RESULT 12

Q8CDRO Q8CDRO PRELIMINARY; PRT; 455 AA.
 AC Q8CDRO;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Transmembrane protease.
 GN TMPSRS5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA THE FANTOM Consortium,
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RT Nature 420:563-573 (2002).
 DR EMBL; AK029714; BAC26577.1; -
 DR MGD; MGI:1933407; Tmprss5.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001354; Peptidase_S1.
 DR InterPro; IPR001190; Srct_receptor.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 SQ SEQUENCE 455 AA; 49669 MW; BE22EB2E7503C74B CRC64;

Query Match 27.0%; Score 631.5; DB 11; Length 455;
 Best Local Similarity 36.2%; Pred. No. 1.8e-52;
 Matches 135; Conservative 63; Mismatches 140; Indels 35; Gaps 10;

QY 79 LDCPLGEDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNWFSACFDNFTEALAEATAC 138
 Db 91 LNCPGVSREELPSLPK--TVSFRNGEDLLLVQVVRAPDMLLVCHEGSGFALGMHIC 148
 QY 139 RQMGYSKPTFRAVEIGDPQDLVDVEITENSQELMRNS-----SGPCLSGSL 186

```

Db      149 KSLGHLRLTQHKAVNL---SDIKL-----NRSQEPAQLSARPGSLVEEAWQSTNCP 201
Qy      187 VSLHCLACGKSLKTRPVVGGEEASVDSNPQVSIQYDKQHVCGGSLDPHVLTAAC-- 244
Db      202 VSLKSCSECGARPLASRIVGGQAVASGRVFWQASVNLGSRHTCGASVLAPHVWVTAACOMY 261
Qy      245 -ERKTDVFNWVRAGSKLGS---FPSLAVAKIIIEFNPMYPKDN---DIALMKLOFP 297
Db      262 SFL--SRSSWRVHAGLVSHGAVRQHQCTWVEKIIP---HPLYSAQNHEDYDVALQLRTP 317
Qy      298 LTFSTGVTRPCLPFFDEELTPATPLWIIIGWGTCKONGKMSDILLQASVQVIDSTR 357
Db      318 INFSDTVGAVCLPAKEQHPQSGCWVSGWGHDTFSTHSSDTLODTWVPLLSLYLNCSS 377
Qy      358 DAYQGEVTEKMKCAGIPEGVDTCQDGGGLMYQS-DQWVGVVSVGMYGCGGPSTPGV 416
Db      378 CMYSALTRMLCAGYLDGRADACQDGGGLVCPSGDTWHLVGVVSWGRGCABENRPGV 437
Qy      417 YTKVSAYLNIWYN 429
Db      438 YAKVAEFLDWIHD 450

```

RESULT 13

```

Q8CJ16 ID Q8CJ16 PRELIMINARY; PRT; 371 AA.
AC Q8CJ16;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenal mitochondrial protease short variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEHD;
RA Omer S., Bicknell A.B., Lowry P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF537099; AN06758.1; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00287; SRCR_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

```

```

Query Match      26.1%; Score 612; DB 11; Length 371;
Best Local Similarity 34.7%; Pred. No. 1.1e-50;
Matches 131; Conservative 62; Mismatches 136; Indels 46; Gaps 11;
Qy      87 EBEVKSFPEGPA-----VAVLSKDRSTLQVLSATGNWFSACFDNTEALA 134
Db      2 EIRCTEEG-GPGIFRAELGDDQQPISFRINGEDLLLEQVVARPDMLLVCHEGWNPALG 60
Qy      135 ETACRQMGYSKFTFAVEIGPDQDLVDVVEITENSQELRMNS-----SGPCL 182

```

```

Db      61 MHICQSLGYFRLTQHKAVNL---SDIKL-----NRSQEPAQLSARPGSLVEEAWQSTNCP 113
Qy      183 SGLVSLHCLACGKSLKTRPVVGGEEASVDSNPQVSIQYDKQHVCGGSLDPHVLTAAC 242
Db      114 SGRIVSLKSECGARPLASRIVGGQAVASGRVFWQASVNLGSRHTCGGSLVAPHVWVTAAC 173
Qy      243 HC---FRKHTDVFNWVRAG---SDKLGFPSLAVAKIIIEFNPMYPKDN---DIALMK 293
Db      174 HCMYSFRL--SRSSWRVHAGLVSHGAVRQHQCTWVEKIIP---HPLYSAQNHEDYDVALQL 229
Qy      294 LQPLPFSGTVPICLPFFDEELTPATPLWIIIGWGTCKONGKMSDILLQASVQVIDSTR 353
Db      230 LRTPIINFSDTVGAVCLPAKEQHPQSGCWVSGWGHDTFSTHSSDTLODTWVPLLSIDL 289
Qy      354 CNADDAYQGEVTEKMKCAGIPEGVDTCQDGGGLMYQS-DQWVGVVSVGMYGCGGSPS 412
Db      290 CNSSCMYSALTRMLCAGYLDGRADACQDGGGLVCPSGDTWHLVGVVSWGRGCABEN 349
Qy      413 TPGVYTKVSAYLNIWYN 429
Db      350 RGVYAKVAEFLDWIHD 366

```

RESULT 14

```

Q8CAN9 ID Q8CAN9 PRELIMINARY; PRT; 777 AA.
AC Q8CAN9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protease (Fragment).
GN PRSS7 OR A130037D21RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK038356; BAC29973.1; --
DR MGD; MGI:1197523; Prss7.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0005044; F:scavenger receptor activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000839; CUB; 2.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR001190; Srcr_receptor.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00629; MAM; 1.
DR Pfam; PF00530; SRCR; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00020; MAMDOMAIN.
DR SMART; SM00042; CUB; 1.
DR SMART; SM00192; LDua; 1.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.

```



```
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 1.
DR PROSITE; PS00740; MAX_1; 1.
DR PROSITE; PS00060; MAX_2; 1.
DR PROSITE; PS02087; SRCR_2; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
FT NON TER 1
SQ SEQUENCE 777 AA; 87314 MW; A18E2F4ECF06D3A8 CRC64;

Query Match 23.3%; Score 545; DB 11; Length 777;
Best Local Similarity 32.5%; Pred. No. 9.8e-44;
Matches 124; Conservative 72; Mismatches 148; Indels 38; Gaps 13;

QY 69 IPRKQCDGELDCPLGEDEEHCVKSFEGPAVAVRLSKDRSTLQVLD-SATGNWFSACFD 127
Db 410 IPLGNLDCSYPHCRDGSDEASCVRP-----LNGTRSNGLVQFNHSHIACAE 459
QY 128 NFEALAEATACRMQGYSSKPTFRAVEI---GPDQDLQDVEITEISQELRMNNSGPGCLSG 184
Db 460 NWTTCISNEVCHLLGLGANSMPISSTGGP-----FVRVYQAPNGSLIILTPSLQCSQD 514
QY 185 SLVSLHC--LACGSLKT-----PRVVGEBASVDSWPQVSIQYDKH-----VCGGSILD 234
Db 515 SLILLQCNHKSCKEKKVTQKSPKIVGGSDAQAGAPWVVALYHRDRSTDRLLCGASLVS 574
QY 235 PHWVLTAAHC-FRKHDTFVNKVRAGSKLGSPFLSAVAKIII--IBFNMPYP---KDND 288
Db 575 SDMLVSAACHVYRNLDFTRTAVLGLHMQSNLTSPOVRRVVDQIVINPHYDRRKVND 634
QY 289 IALMKLQPLFTSGTRVPCILPFDEELTPATPLWIGWFTKONGKMSDILLOASVQV 348
Db 635 IAMEHLEPKVNYTDYIQPICLPEENQIFPGRTCSAGWGYDKINGSTV-DVLKXADVPL 693
QY 349 IDSTRCNADDAVOGEVEKMKWAGIPEGGVDTCCGDSGGLMYO--SDQHVVGIVSWGYG 407
Db 694 ISNEKCO-QQLPEYNITSMICAGVEGGIDSCGDSGGLMCGENNRWFLVGVTSFGVQ 752
QY 408 CGGPSTPGVYTKVSAYLNWIYN 429
Db 753 CALPNHPGVYVRVSOQFIEWHS 774

RESULT 15
Q8VHJ4 PRELIMINARY; PRT; 417 AA.
AC Q8VHJ4;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Airway trypsin-like protease.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
ZN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Trachea;
RA Hansen I.A.; Fasnacht M.; Hammer F.; Schammann M.; Allolio B.;
RT "Cloning and characterization of RAT, the airway trypsin-like protease
of Rattus norvegicus.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AF453776; AAL50817.1; -.
DR HSSP; P00761; IANI.
DR MEROPS; S01.047; -.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase S1.
```

```
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR SMART; SM00200; SEA; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00024; SEA; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 417 AA; 46287 MW; DB9504158B018E21 CRC64;

Query Match 22.9%; Score 536; DB 11; Length 417;
Best Local Similarity 34.4%; Pred. No. 3.1e-43;
Matches 129; Conservative 48; Mismatches 98; Indels 100; Gaps 14;

QY 89 HCYKSPFEGPAVAV-----RLSKD-----RSTLQVLDSATGNWFSACFDNFEA 132
Db 101 HVVKLRKEGSGVVADVVMKFRSSKRNKKAIKTRIQSVLQRL-SSSGN----- 147
QY 133 LAETACRMQGYSSKPTFRAVEIGP-----DQDLQDVEITEISQELRMNNSGPGCL 182
Db 148 -----CGARPDLITLSEERIIGCTQAEIGTGMQVLSLQNNVHCHGGLISNLW 220
QY 183 SGLVSLHCLACGK-----SLKTPRVVGGEASVDSWPQVSIQYDKH-----VCGGSILD 237
Db 172 -----CGARPDLITLSEERIIGCTQAEIGTGMQVLSLQNNVHCHGGLISNLW 220
QY 238 VLTAAHCFRKHDTFVNKVRAGSKLGSPFLSAVAKIII--IBFNMPYKNDIALMKL 294
Db 221 VLTAAHCFRYSNPQOQTATFGVSTIS--PRLRVRVRAILAHAEYNSI-TRDNIAVQL 277
QY 295 QPPLTFSGTVRPICLPFDEELTPATPLWIGWFTKONGKMSDILLOASVQVIDSTRC 354
Db 278 DRVTTFTRIHRVCLPAAQCNIMPDSVAVYTWGSLTYGNTVTN-LQGEVRIVSSEVC 336
QY 355 NADDAVOGEVEKMKWAGIPEGGVDTCCGDSGGLMYQSDQ--WHVVGIVSWGYGCGGPS 412
Db 337 NEPAGYGGSVLPGLMCGAVRSGAVDACQDGGGLPQVEDTRRLRFVVGIVSWGYGCGGLN 396
QY 413 TPGVYTKVSAYLNWI 427
Db 397 KPGVYTRVTAAYRNWI 411

Search completed: June 7, 2004, 08:42:05
Job time : 48 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 08:41:12 ; Search time 48 Seconds
(without alignments)
2549.632 Million cell updates/sec

Title: US-10-030-688-2

Perfect score: 2342
Sequence: 1 MDPDSQPLNSLDVPLRKP.....VYTKVSAYLNWYVWKAEL 435

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/FCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/FCRUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2338	99.8	435	12	US-09-968-415-6
2	2338	99.8	435	14	Sequence 6, Appli
3	2337	99.8	437	9	US-10-180-719-6
4	2337	99.8	437	9	Sequence 6, Appli
5	2337	99.8	437	15	US-09-851-588-8
6	2337	99.8	437	15	Sequence 8, Appli
7	2337	99.8	437	15	Sequence 779, App
8	2337	99.8	437	15	Sequence 791, App
9	2337	99.8	437	15	Sequence 831, App
10	2337	99.8	437	15	Sequence 89, Appl
11	2337	99.8	437	15	Sequence 89, Appl
12	2337	99.8	437	15	Sequence 4, Appli
13	2337	99.8	437	15	Sequence 4, Appli
14	2337	99.8	437	15	Sequence 7, Appli
15	2337	99.8	437	15	Sequence 275, App
16	2337	99.8	437	15	Sequence 330, App
17	2337	99.8	437	15	Sequence 330, App
18	2337	99.8	437	15	Sequence 330, App

16 2297.5 98.1 432 12 US-10-205-890-330
17 2297.5 98.1 432 12 US-10-208-024-330
18 2297.5 98.1 432 12 US-10-201-853-330
19 2297.5 98.1 432 12 US-10-063-745-112
20 2297.5 98.1 432 12 US-10-063-512-112
21 2297.5 98.1 432 12 US-10-063-513-112
22 2297.5 98.1 432 12 US-10-063-515-112
23 2297.5 98.1 432 12 US-10-063-549-112
24 2297.5 98.1 432 12 US-10-063-569-112
25 2297.5 98.1 432 12 US-10-063-551-112
26 2297.5 98.1 432 12 US-10-174-581-330
27 2297.5 98.1 432 12 US-10-176-483-330
28 2297.5 98.1 432 12 US-10-176-749-330
29 2297.5 98.1 432 12 US-10-176-914-330
30 2297.5 98.1 432 12 US-10-176-915-330
31 2297.5 98.1 432 12 US-10-006-485A-275
32 2297.5 98.1 432 12 US-10-013-907A-275
33 2297.5 98.1 432 12 US-10-015-499A-275
34 2297.5 98.1 432 12 US-10-063-555-112
35 2297.5 98.1 432 12 US-10-063-563-112
36 2297.5 98.1 432 12 US-10-063-594-112
37 2297.5 98.1 432 12 US-10-063-553-112
38 2297.5 98.1 432 12 US-10-063-554-112
39 2297.5 98.1 432 12 US-10-176-484-330
40 2297.5 98.1 432 12 US-10-180-550-330
41 2297.5 98.1 432 12 US-10-183-014-330
42 2297.5 98.1 432 12 US-10-187-738-330
43 2297.5 98.1 432 12 US-10-187-740-330
44 2297.5 98.1 432 12 US-10-187-883-330
45 2297.5 98.1 432 12 US-10-194-363-330

ALIGNMENTS

RESULT 1
US-09-968-415-6
; Sequence 6, Application US/09968415
; Publication No. US20020086334A1

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/968,415

FILING DATE: 26-Sep-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/659,151

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheila

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

```

;
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-09-968-415-6

Query Match          99.8%; Score 2338; DB 12; Length 435;
Best Local Similarity 99.8%; Pred. No. 1.2e-226;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60
DB 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60
QY 61 LCGQPLHFI PRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
DB 61 LCGQPLHFI PRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNTEALAEACACRGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNNSGP 180
DB 121 WFSACFDNTEALAEACACRGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNNSGP 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
QY 241 AAHCFRKHDTVENKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
DB 241 AAHCFRKHDTVENKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAY 360
DB 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAY 360
QY 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVC GGPSTPGVYTKV 420
DB 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVC GGPSTPGVYTKV 420

```

RESULT 2

```

US-10-180-719-6
; Sequence 6, Application US/10180719
; Publication No. US20030166246A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hillman, Jennifer L.
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; Tang, Tom Y.
; Shah, Purvi
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

```

;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/180,719
; FILING DATE: 25-Jun-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,271
; FILING DATE: 16-Jan-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Mohan-Peterson, Sheela
; REGISTRATION NUMBER: 41,201
; REFERENCE/DOCKET NUMBER: PF-0458 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: COLNNOT13
; CLONE: 1337018
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :
US-10-180-719-6

Query Match          99.8%; Score 2338; DB 14; Length 435;
Best Local Similarity 99.8%; Pred. No. 1.2e-226;
Matches 434; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60
DB 1 MDPDSQPLNSLDVKPLKPRIPMETFRKVGIPPIIIALLSLASIIIVVLIKVILDKYIF 60
QY 61 LCGQPLHFI PRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
DB 61 LCGQPLHFI PRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGN 120
QY 121 WFSACFDNTEALAEACACRGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNNSGP 180
DB 121 WFSACFDNTEALAEACACRGYSSKPTFRAVEIGPDQDLDVVEITENSQELMRNNSGP 180
QY 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
DB 181 CLSGSLVSLHCLACGSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWLT 240
QY 241 AAHCFRKHDTVENKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
DB 241 AAHCFRKHDTVENKVRAGSKLGSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTF 300
QY 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAY 360
DB 301 SGTVRPICLPFFDEELTPATPLMIIGWFTKQNGKMSDILLQASVQVIDSTRCNAD DAY 360
QY 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVC GGPSTPGVYTKV 420
DB 361 QGEVTEKMCAGIPEGVDTCQDSCGGLMYQSDQHVHVGIVSWGVC GGPSTPGVYTKV 420
QY 421 SAYLNWIYNNVWKAEL 435
DB 421 SAYLNWIYNNVWKAEL 435

```

RESULT 3

```

US-09-851-588-8
; Sequence 8, Application US/09851588
; Patent No. US20020042067A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.

```

; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
; TITLE OF INVENTION: OF SCREENING FOR COLORECTAL CANCER MODULATORS

; FILE REFERENCE: A-68929-1/DJB/JUD/AWS

; CURRENT APPLICATION NUMBER: US/09/851,588

; CURRENT FILING DATE: 2001-09-24

; PRIOR APPLICATION NUMBER: US/09/642,252

; PRIOR FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: US/09/656,002

; PRIOR FILING DATE: 2000-09-06

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 8

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-851-588-8

Query Match 99.8%; Score 2337; DB 9; Length 437;

Best Local Similarity 100.0%; Pred. No. 1.6e-226;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 DPDSQPLNSLDVVKPRIPMETFRKVGPIIITALLSLASIIIVVLLIKVILDKYFL 61

Db 4 DPDSQPLNSLDVVKPRIPMETFRKVGPIIITALLSLASIIIVVLLIKVILDKYFL 63

Qy 62 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 121

Db 64 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 123

Qy 122 FSACFDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 181

Db 124 FSACFDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 183

Qy 182 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWPNQVSIQYDKQHVCGSILDPHWLTA 241

Db 184 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWPNQVSIQYDKQHVCGSILDPHWLTA 243

Qy 242 AHCERKHTDVFNNKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301

Db 244 AHCERKHTDVFNNKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303

Qy 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQY 361

Db 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQY 363

Qy 362 GEVTEKMCAGIPEGGVDTCCQSDSGPLMYQSDQHWVGVISWGYCGGPGSTPGVYTKVS 421

Db 364 GEVTEKMCAGIPEGGVDTCCQSDSGPLMYQSDQHWVGVISWGYCGGPGSTPGVYTKVS 423

Qy 422 AYLNWIYNNVKAEL 435

Db 424 AYLNWIYNNVKAEL 437

RESULT 4

US-10-295-027-779

; Sequence 779, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

RESULT 5

US-10-295-027-791

; Sequence 791, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 779

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-295-027-779

Query Match

Best Local Similarity

Matches 434; Conservative

99.8%; Score 2337; DB 15; Length 437;

100.0%; Pred. No. 1.6e-226;

Mismatches 0; Indels 0; Gaps 0;

Qy 2 DPDSQPLNSLDVVKPRIPMETFRKVGPIIITALLSLASIIIVVLLIKVILDKYFL 61

Db 4 DPDSQPLNSLDVVKPRIPMETFRKVGPIIITALLSLASIIIVVLLIKVILDKYFL 63

Qy 62 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 121

Db 64 CGQPLHFIIPRKQCDGELDCPLGEDEEHCVKSPPEGPAVAVLSKDRSTLQVLD SATGNW 123

Qy 122 FSACFDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 181

Db 124 FSACFDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLVDVEITENSQELMRNSSGPC 183

Qy 182 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWPNQVSIQYDKQHVCGSILDPHWLTA 241

Db 184 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWPNQVSIQYDKQHVCGSILDPHWLTA 243

Qy 242 AHCERKHTDVFNNKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301

Db 244 AHCERKHTDVFNNKVRAGSKLGPSPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303

Qy 302 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQY 361

Db 304 GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLOASVQVIDSTRCNADDAQY 363

Qy 362 GEVTEKMCAGIPEGGVDTCCQSDSGPLMYQSDQHWVGVISWGYCGGPGSTPGVYTKVS 421

Db 364 GEVTEKMCAGIPEGGVDTCCQSDSGPLMYQSDQHWVGVISWGYCGGPGSTPGVYTKVS 423

Qy 422 AYLNWIYNNVKAEL 435

Db 424 AYLNWIYNNVKAEL 437

APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 791
LENGTH: 437
TYPE: PRT
ORGANISM: Homo sapiens
US-10-295-027-791

Query Match 99.8%; Score 2337; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.6e-226; Mismatches 0; Indels 0; Gaps 0;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLKDYFL 61
DB 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLKDYFL 63
QY 62 CGQPLHFIPRKQICDGLDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLOVLDSATGNW 121
DB 64 CGQPLHFIPRKQICDGLDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNMSGPC 181
DB 124 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNMSGPC 183
QY 182 LSGSLVSLHCLACGSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
DB 184 LSGSLVSLHCLACGSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 243
QY 242 AHCFRKHTDVFNKVRAGSKLSPSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS 301
DB 244 AHCFRKHTDVFNKVRAGSKLSPSPSLAVAKIIIEFNPMYPKNDIALMKLQPLTFS 303
QY 302 GTVRPCLPFFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQVLDTRCNADDAAY 361
DB 304 GTVRPCLPFFDEELTPATPLMIIGWFTKQNGKMSDILLOASVQVLDTRCNADDAAY 363
QY 362 GEVTEKQWCAGIPEGGVDTCCQDGGGLMYQSDQWHVGVISWGYCGGPGSTPGVYTKVS 421

DB 364 GEVTEKQWCAGIPEGGVDTCCQDGGGLMYQSDQWHVGVISWGYCGGPGSTPGVYTKVS 423
QY 422 AYLNWYNNWKAEL 435
DB 424 AYLNWYNNWKAEL 437
RESULT 6
US-10-295-027-831
Sequence 831, Application US/10295027
Publication No. US20030232350A1
GENERAL INFORMATION:
APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Ginsberg, Wendy M.
APPLICANT: Gish, Kurt C.
APPLICANT: Glynn, Richard
APPLICANT: Hevezi, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT APPLICATION NUMBER: US/10/295,027
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 831
LENGTH: 437
TYPE: PRT
ORGANISM: Homo sapiens
US-10-295-027-831

Query Match 99.8%; Score 2337; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.6e-226; Mismatches 0; Indels 0; Gaps 0;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLKDYFL 61
DB 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVLKDYFL 63
QY 62 CGQPLHFIPRKQICDGLDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLOVLDSATGNW 121
DB 64 CGQPLHFIPRKQICDGLDCPLGEDEEHCVKSPFPGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNMSGPC 181
DB 124 FSACFNFTEALAEACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNMSGPC 183
QY 182 LSGSLVSLHCLACGSLKTPRVVVGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241

Db 184 LSGSLVSLHCLACGSLKTPRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDHPHWLTA 243
QY 242 AHCPRKHTDVFNKVRAGSKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 244 AHCPRKHTDVFNKVRAGSKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
QY 302 GTVRPCLPFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTRCNADDAYQ 361
Db 304 GTVRPCLPFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTRCNADDAYQ 363
QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGLPMYQSDQHHVVGIVSWGCGGSPSTPGVYTKVS 421
Db 364 GEVTEKMCAGIPEGGVDTCCGDSGGLPMYQSDQHHVVGIVSWGCGGSPSTPGVYTKVS 423
QY 422 AYLNWIYNWKAEL 435
Db 424 AYLNWIYNWKAEL 437

RESULT 7

US-10-295-027-1196
; Sequence 1196, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1196
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1196

Query Match 99.8%; Score 2337; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 1,6e-226;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPIMPETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 61
Db 4 DPDSQPLNSLDVKPLRKPIMPETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGOPLHFIIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 121
Db 64 CGOPLHFIIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACFNFTALAEATACRQWYSSKPTFRAVEIGPQDLDVVEITENSQELVRNMSGPC 181
Db 124 FSACFNFTALAEATACRQWYSSKPTFRAVEIGPQDLDVVEITENSQELVRNMSGPC 183
QY 182 LSGSLVSLHCLACGSKLKTFRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDHPHWLTA 241
Db 184 LSGSLVSLHCLACGSKLKTFRVVGGEASVDSWPMQVSIQYDKQHVCGGSLDHPHWLTA 243
QY 242 AHCPRKHTDVFNKVRAGSKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 301
Db 244 AHCPRKHTDVFNKVRAGSKLGSPPSLAVAKIIIEFNPMYPKNDIALMKLOPPLTFS 303
QY 302 GTVRPCLPFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTRCNADDAYQ 361
Db 304 GTVRPCLPFDEELTPATPLWIIIGWFTKQNGGKMSDILLQASVOVIDSTRCNADDAYQ 363
QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGLPMYQSDQHHVVGIVSWGCGGSPSTPGVYTKVS 421
Db 364 GEVTEKMCAGIPEGGVDTCCGDSGGLPMYQSDQHHVVGIVSWGCGGSPSTPGVYTKVS 423
QY 422 AYLNWIYNWKAEL 435
Db 424 AYLNWIYNWKAEL 437

RESULT 8

US-10-173-999-89
; Sequence 89, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 89
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-89

Query Match 99.8%; Score 2337; DB 15; Length 437;
Best Local Similarity 100.0%; Pred. No. 1,6e-226;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPIMPETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 61
Db 4 DPDSQPLNSLDVKPLRKPIMPETFRKVGIPITIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGOPLHFIIPRKQLCDGELDCPLGEDDEHCVKSPPEGPAVAVRLSKDRSTLOVLDSATGNW 121

Db 64 CGQPLHPIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPG 181
 Db 124 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPG 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 243
 QY 242 AHCFRKHDTVFNKVRAGSKLGFSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
 Db 244 AHCFRKHDTVFNKVRAGSKLGFSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
 QY 302 GTVRPCLPFDFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 361
 Db 304 GTVRPCLPFDFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 363
 QY 362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGPGSTPGVYTKVS 421
 Db 364 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGPGSTPGVYTKVS 423
 QY 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 9

US-09-776-191-4
 ; Sequence 4, Application US/09776191
 ; Publication No. US20030119168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edwin L. Madison
 ; APPLICANT: Edgar O. Ong
 ; APPLICANT: Jiunn-Chern Yeh
 ; APPLICANT: Corvas International, Inc.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
 ; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
 ; FILE REFERENCE: 24745-1607
 ; CURRENT APPLICATION NUMBER: US/09/776,191
 ; CURRENT FILING DATE: 2001-02-02
 ; PRIOR APPLICATION NUMBER: 60/213,124
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 60/234,840
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 60/179,982
 ; PRIOR FILING DATE: 2000-02-03
 ; PRIOR APPLICATION NUMBER: 60/183,542
 ; PRIOR FILING DATE: 2000-02-18
 ; PRIOR APPLICATION NUMBER: 09/657,968
 ; PRIOR FILING DATE: 2000-02-08
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-776-191-4

Query Match 99.0%; Score 2319; DB 10; Length 437;
 Best Local Similarity 99.5%; Pred. No. 1e-224;
 Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 61
 Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 63
 QY 62 CGQPLHPIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 121
 Db 64 CGQPLHPIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPG 181

Db 124 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPG 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 243
 QY 242 AHCFRKHDTVFNKVRAGSKLGFSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
 Db 244 AHCFRKHDTVFNKVRAGSKLGFSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
 QY 302 GTVRPCLPFDFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 361
 Db 304 GTVRPCLPFDFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 363
 QY 362 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGPGSTPGVYTKVS 421
 Db 364 GEVTERKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGVCYCGGPGSTPGVYTKVS 423
 QY 422 AYLNWIYNNWKAEL 435
 Db 424 AYLNWIYNNWKAEL 437

RESULT 10

US-10-156-214A-4
 ; Sequence 4, Application US/10156214A
 ; Publication No. US2004001801A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edwin L. Madison
 ; APPLICANT: Joseph Edward Sempke
 ; APPLICANT: George P. Vlasuk
 ; APPLICANT: Scott Jeffrey Kemp
 ; APPLICANT: Mallareddy Komandla
 ; APPLICANT: Daniel Vanna Siev
 ; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic Use
 ; TITLE OF INVENTION: Thereof
 ; FILE REFERENCE: 24745-1611
 ; CURRENT APPLICATION NUMBER: US/10/156,214A
 ; CURRENT FILING DATE: 2002-05-23
 ; NUMBER OF SEQ ID NOS: 611
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 437
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-156-214A-4

Query Match 99.0%; Score 2319; DB 15; Length 437;
 Best Local Similarity 99.5%; Pred. No. 1e-224; 2; Indels 0; Gaps 0;
 Matches 432; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 61
 Db 4 DPDSQPLNSLDVKPLRKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYFL 63
 QY 62 CGQPLHPIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 121
 Db 64 CGQPLHPIPRKQLCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLQVLD SATGNW 123
 QY 122 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPG 181
 Db 124 FSACFNFTALAEATACRQMGYSKPTFRAVEIGPDQDLVDVEITENSQELMRNNSGPG 183
 QY 182 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 241
 Db 184 LSGSLVSLHCLACGKSLKTPRVVGGSEASVDSWPQVSIQYDKQHVCGGSIIDPHWVLT 243
 QY 242 AHCFRKHDTVFNKVRAGSKLGFSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 301
 Db 244 AHCFRKHDTVFNKVRAGSKLGFSPSLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS 303
 QY 302 GTVRPCLPFDFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAVQ 361

Db 304 GTVRLICLPFFDEELTPATPLWIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQ 363
QY 362 GEVTERKMCAGIPGEGVDTCCGDSGGLMYOSDQHHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 364 GEVTERKMCAGIPGEGVDTCCGDSGGLMYOSDQHHVVGIVSWGCGGPGSTPGVYTKVS 423
QY 422 AYLNWIYNNWKAE 435
Db 424 AYLNWIYNNWKAE 437

RESULT 11

US-09-888-257A-7

; Sequence 7, Application US/09888257A

; Publication No. US20030060612A1

; GENERAL INFORMATION:

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Polakis, Paul

; APPLICANT: Smith, Victoria

; APPLICANT: Wood, William I.

; APPLICANT: Wu, Thomas D.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

; TREATMENT OF TUMOR

; FILE REFERENCE: P5002R1

; CURRENT APPLICATION NUMBER: US/09/888,257A

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/063,540

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: US 60/089,653

; PRIOR FILING DATE: 1998-06-17

; PRIOR APPLICATION NUMBER: US 60/099,792

; PRIOR FILING DATE: 1998-09-10

; PRIOR APPLICATION NUMBER: US 60/103,678

; PRIOR FILING DATE: 1998-10-08

; PRIOR APPLICATION NUMBER: US 60/235,451

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: PCT/US99/12252

; PRIOR FILING DATE: 1999-08-02

; PRIOR APPLICATION NUMBER: PCT/US99/20111

; PRIOR FILING DATE: 1999-09-01

; PRIOR APPLICATION NUMBER: PCT/US00/04342

; PRIOR FILING DATE: 2000-02-18

; PRIOR APPLICATION NUMBER: PCT/US00/05841

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: PCT/US00/08439

; PRIOR FILING DATE: 2000-03-30

; PRIOR APPLICATION NUMBER: PCT/US00/23328

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: PCT/US00/32678

; PRIOR FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: PCT/US01/06520

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: PCT/US01/06666

; PRIOR FILING DATE: 2001-03-01

; NUMBER OF SEQ ID NOS: 10

; SEQ ID NO 7

; LENGTH: 432

; TYPE: PRT

; ORGANISM: Homo Sapien

US-09-888-257A-7

Query Match

Best Local Similarity 98.1%; Score 2297.5; DB 10; Length 432;

Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSQPLNSLDVKLRPRIPMETFRKVGIPITIALSLASIIIVWLKIVLDKYYFL 61

Db 4 DPDSQPLNSLDVKLRPRIPMETFRKVGIPITIALSLASIIIVWLKIVLDKYYFL 63

QY 62 CGSPLHEIPEKQCDGELDCPLGEDBEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 121
Db 64 CGSPLHEIPEKQCDGELDCPLGEDBEHCVKSPPEGPAVAVRLSKDRSTLQVLDSATGNW 123
QY 122 FSACFDFNTEALAEATACRQMGYSKPTFRAVEIGPDQDLDDVVEITENSQELRMNMSGPC 181
Db 124 FSACFDFNTEALAEATACRQMGYS----RAVEIGPDQDLDDVVEITENSQELRMNMSGPC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSLDHPHWLTA 241
Db 179 LSGSLVSLHCLACGKSLKTPRVVVGGEASVDSWPMQVSIQYDKQHVCGGSLDHPHWLTA 238
QY 242 AHCFRKHTDVFNWVKVAGSDKLGSPFSLAVAKIIIEFNPMYKNDIALMKIQFPLTFS 301
Db 239 AHCFRKHTDVFNWVKVAGSDKLGSPFSLAVAKIIIEFNPMYKNDIALMKIQFPLTFS 298
QY 302 GTVRPICLPFFDEELTPATPLWIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQ 361
Db 299 GTVRPICLPFFDEELTPATPLWIGWFTKQNGKMSDILLQASVQVIDSTRCNADDAQ 358
QY 362 GEVTERKMCAGIPGEGVDTCCGDSGGLMYOSDQHHVVGIVSWGCGGPGSTPGVYTKVS 421
Db 359 GEVTERKMCAGIPGEGVDTCCGDSGGLMYOSDQHHVVGIVSWGCGGPGSTPGVYTKVS 418
QY 422 AYLNWIYNNWKAE 435
Db 419 AYLNWIYNNWKAE 432

RESULT 12

US-09-946-374-275

; Sequence 275, Application US/09946374

; Publication No. US20030073129A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Deenoyers, Luc

; APPLICANT: Eaton, Dan L.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

; APPLICANT: Roy, Margaret Ann

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Williams, P. Mickey

; APPLICANT: Wood, William I.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; ACIDS Encoding the Same

; FILE REFERENCE: P2830P1C1

; CURRENT APPLICATION NUMBER: US/09/946,374

; CURRENT FILING DATE: 2001-09-04

; PRIOR APPLICATION NUMBER: 60/098716

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098723

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098749

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098750

; PRIOR FILING DATE: 1998-09-01

; PRIOR APPLICATION NUMBER: 60/098803

; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/098821

; PRIOR FILING DATE: 1998-09-02

1	PRIOR FILING DATE: 1998-09-23	
2	PRIOR APPLICATION NUMBER: 60/101475	
3	PRIOR FILING DATE: 1998-09-23	
4	PRIOR APPLICATION NUMBER: 60/101476	
5	PRIOR FILING DATE: 1998-09-23	
6	PRIOR APPLICATION NUMBER: 60/101477	
7	PRIOR FILING DATE: 1998-09-23	
8	PRIOR APPLICATION NUMBER: 60/101479	
9	PRIOR FILING DATE: 1998-09-23	
10	PRIOR APPLICATION NUMBER: 60/101738	
11	PRIOR FILING DATE: 1998-09-24	
12	PRIOR APPLICATION NUMBER: 60/101741	
13	PRIOR FILING DATE: 1998-09-24	
14	PRIOR APPLICATION NUMBER: 60/101743	
15	PRIOR FILING DATE: 1998-09-24	
16	PRIOR APPLICATION NUMBER: 60/101915	
17	PRIOR FILING DATE: 1998-09-24	
18	PRIOR APPLICATION NUMBER: 60/101916	
19	PRIOR FILING DATE: 1998-09-24	
20	PRIOR APPLICATION NUMBER: 60/102207	
21	PRIOR FILING DATE: 1998-09-29	
22	PRIOR APPLICATION NUMBER: 60/102240	
23	PRIOR FILING DATE: 1998-09-29	
24	PRIOR APPLICATION NUMBER: 60/102330	
25	PRIOR FILING DATE: 1998-09-29	
26	PRIOR APPLICATION NUMBER: 60/102331	
27	PRIOR FILING DATE: 1998-09-29	
28	PRIOR APPLICATION NUMBER: 60/102484	
29	PRIOR FILING DATE: 1998-09-30	
30	PRIOR APPLICATION NUMBER: 60/102487	
31	PRIOR FILING DATE: 1998-09-30	
32	PRIOR APPLICATION NUMBER: 60/102570	
33	PRIOR FILING DATE: 1998-09-30	
34	PRIOR APPLICATION NUMBER: 60/102571	
35	PRIOR FILING DATE: 1998-09-30	
36	PRIOR APPLICATION NUMBER: 60/102684	
37	PRIOR FILING DATE: 1998-10-01	
38	PRIOR APPLICATION NUMBER: 60/102687	
39	PRIOR FILING DATE: 1998-10-01	
40	PRIOR APPLICATION NUMBER: 60/102965	
41	PRIOR FILING DATE: 1998-10-02	
42	PRIOR APPLICATION NUMBER: 60/103258	
43	PRIOR FILING DATE: 1998-10-06	
44	PRIOR APPLICATION NUMBER: 60/103314	
45	PRIOR FILING DATE: 1998-10-07	
46	PRIOR APPLICATION NUMBER: 60/103315	
47	PRIOR FILING DATE: 1998-10-07	
48	PRIOR APPLICATION NUMBER: 60/103328	
49	PRIOR FILING DATE: 1998-10-07	
50	PRIOR APPLICATION NUMBER: 60/103395	
51	PRIOR FILING DATE: 1998-10-07	
52	PRIOR APPLICATION NUMBER: 60/103396	
53	PRIOR FILING DATE: 1998-10-07	
54	PRIOR APPLICATION NUMBER: 60/103401	
55	PRIOR FILING DATE: 1998-10-07	
56	PRIOR APPLICATION NUMBER: 60/103449	
57	PRIOR FILING DATE: 1998-10-06	
58	PRIOR APPLICATION NUMBER: 60/103633	
59	PRIOR FILING DATE: 1998-10-08	
60	PRIOR APPLICATION NUMBER: 60/103678	
61	PRIOR FILING DATE: 1998-10-08	
62	PRIOR APPLICATION NUMBER: 60/103679	
63	PRIOR FILING DATE: 1998-10-08	
64	PRIOR APPLICATION NUMBER: 60/103711	
65	PRIOR FILING DATE: 1998-10-08	
66	PRIOR APPLICATION NUMBER: 60/104257	
67	PRIOR FILING DATE: 1998-10-14	
68	PRIOR APPLICATION NUMBER: 60/104987	
69	PRIOR FILING DATE: 1998-10-20	
70	PRIOR APPLICATION NUMBER: 60/105000	
71	PRIOR FILING DATE: 1998-10-20	

; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 98.1%; Score 2297.5; DB 10; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222;
Matches 429; Conservative 0; Mismatches 5; Gaps 1;

QY 2 DPDSQPLNSLDVFKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVFKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGQPLHFIPIRQKQCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 121
DB 64 CGQPLHFIPIRQKQCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELRMNSSGPC 181
DB 124 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELRMNSSGPC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWWLTA 241
DB 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWWLTA 238
QY 242 AHCPRKHTDVFNWVKRAGSKLGSPSLAVAKIIIEFNPMYKNDIALMKLOFPLTFS 301
DB 239 AHCPRKHTDVFNWVKRAGSKLGSPSLAVAKIIIEFNPMYKNDIALMKLOFPLTFS 298
QY 302 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKVSILLQASVQVIDSTRCNADDAAYQ 361
DB 299 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKVSILLQASVQVIDSTRCNADDAAYQ 358
QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 421
DB 359 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 418
QY 422 AYLNWYNNWKAEL 435
DB 419 AYLNWYNNWKAEL 432

RESULT 13
US-10-206-915-330
; Sequence 330, Application US/10206915
; Publication No. US20040029221A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1CS3
; CURRENT APPLICATION NUMBER: US/10/206,915
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586

; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-206-915-330

Query Match 98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222;
Matches 429; Conservative 0; Mismatches 5; Gaps 1;

QY 2 DPDSQPLNSLDVFKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVLIDKYFL 61
DB 4 DPDSQPLNSLDVFKPRIPMETFRKVGPIIIALLSLASIIIVVLIKVLIDKYFL 63
QY 62 CGQPLHFIPIRQKQCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 121
DB 64 CGQPLHFIPIRQKQCDGELDCPLGEDEEHCVKSPFEGPAVAVRLSKDRSTLOVLDSATGNW 123
QY 122 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELRMNSSGPC 181
DB 124 FSACPDNTEALAEATACROMGYSSKPTFRAVEIGPDQDLDVVEITENSQELRMNSSGPC 178
QY 182 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWWLTA 241
DB 179 LSGSLVSLHCLACGKSLKTPRVVGGEEASVDSWPQVSIQYDKQHVCGGSLDPHWWLTA 238
QY 242 AHCPRKHTDVFNWVKRAGSKLGSPSLAVAKIIIEFNPMYKNDIALMKLOFPLTFS 301
DB 239 AHCPRKHTDVFNWVKRAGSKLGSPSLAVAKIIIEFNPMYKNDIALMKLOFPLTFS 298
QY 302 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKVSILLQASVQVIDSTRCNADDAAYQ 361
DB 299 GTVRPCLPFFDEELTPATPLWIIIGWFTKQNGKVSILLQASVQVIDSTRCNADDAAYQ 358
QY 362 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 421
DB 359 GEVTEKMCAGIPEGGVDTCCGDSGGPLMYQSDQHVHVGIVSWGCGGPGSTPGVYTKVS 418
QY 422 AYLNWYNNWKAEL 435
DB 419 AYLNWYNNWKAEL 432

RESULT 14
US-10-199-670-330
; Sequence 330, Application US/10199670
; Publication No. US20040033560A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C401
; CURRENT APPLICATION NUMBER: US/10/199,670
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-199-670-330

Query Match          98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSOPLNSLDVVKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 61
DB 4 DPDSOPLNSLDVVKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 63

QY 62 CGOPLHFI PRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
DB 64 CGOPLHFI PRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELRMNRS SGP 181
DB 124 FSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELRMNRS SGP 178

QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
DB 179 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 238

QY 242 AHCFRKHTDVFNVKVRAGSKLGSFPSPSLAVAKIIIEFNPMYKNDIALMKLOPPLTFS 301
DB 239 AHCFRKHTDVFNVKVRAGSKLGSFPSPSLAVAKIIIEFNPMYKNDIALMKLOPPLTFS 298

QY 302 GTVRICLPFFDEELTPATPLMIIGWGTCKONGKMSDILLQASVQVLDSTRCNADDA YQ 361
DB 299 GTVRICLPFFDEELTPATPLMIIGWGTCKONGKMSDILLQASVQVLDSTRCNADDA YQ 358

QY 362 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGCGGPGTGPVYTKVS 421
DB 359 GEVTEKMCAGIPEGGVDTCCQDSGGPLMYQSDQHVHVGIVSWGCGGPGTGPVYTKVS 418

QY 422 AYLNWIYNVWKAE 435
```

```

DB 419 AYLNWIYNVWKAE 432

RESULT 15
US-10-201-858-330
; Sequence 330, Application US/10201858
; Publication No. US20040038337A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C464
; CURRENT APPLICATION NUMBER: US/10/201,858
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 330
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-201-858-330

Query Match          98.1%; Score 2297.5; DB 12; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.5e-222;
Matches 429; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 2 DPDSOPLNSLDVVKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 61
DB 4 DPDSOPLNSLDVVKPRIPMETFRKVGIPITIIALLSLASIIIVVLIKVILDKYYFL 63

QY 62 CGOPLHFI PRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 121
DB 64 CGOPLHFI PRKQLCDGELDCPLGEDEEHCVKSFPEGPAVAVRLSKDRSTLQVLD SATGNW 123

QY 122 FSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELRMNRS SGP 181
DB 124 FSACFDNFTALAEACROMGYSSKPTFRAVEIGPDQDLDDVVEITENSQELRMNRS SGP 178

QY 182 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 241
DB 179 LSGSLVSLHCLACGSKLTPRVVGGEEASVDSWPQVSIQYDKQHVCGSILDPHWLTA 238
```

QY	242	AHCFRKHTDVFNKVRAGSKLGSFPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS	301
Db	239	AHCFRKHTDVFNKVRAGSKLGSFPLAVAKIIIEFNPMYPKNDIALMKLQFPLTFS	298
QY	302	GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQ	361
Db	299	GTVRPICLPFFDEELTPATPLWIIIGWFTKQNGKMSDILLQASVQVIDSTRCNADDDAYQ	358
QY	362	GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVGVIVSWGYCGGSPSTPGVYTKVS	421
Db	359	GEVTERKMCAGIPEGGVDTCCGDSGGPLMYQSDQWHVGVIVSWGYCGGSPSTPGVYTKVS	418
QY	422	AYLNWIYNVWKAEL	435
Db	419	AYLNWIYNVWKAEL	432

Search completed: June 7, 2004, 08:47:29
Job time : 49 secs

This Page Blank (uspto)